IR-NLP Talk #3

Relation Extraction for Knowledge Base Completion

Bayu Distiawan Trisedya

Gerhard Weikum

Jianzhong Qi

Rui Zhang

THE UNIVERSITY OF MELBOURNE

ACL 2019

Florence

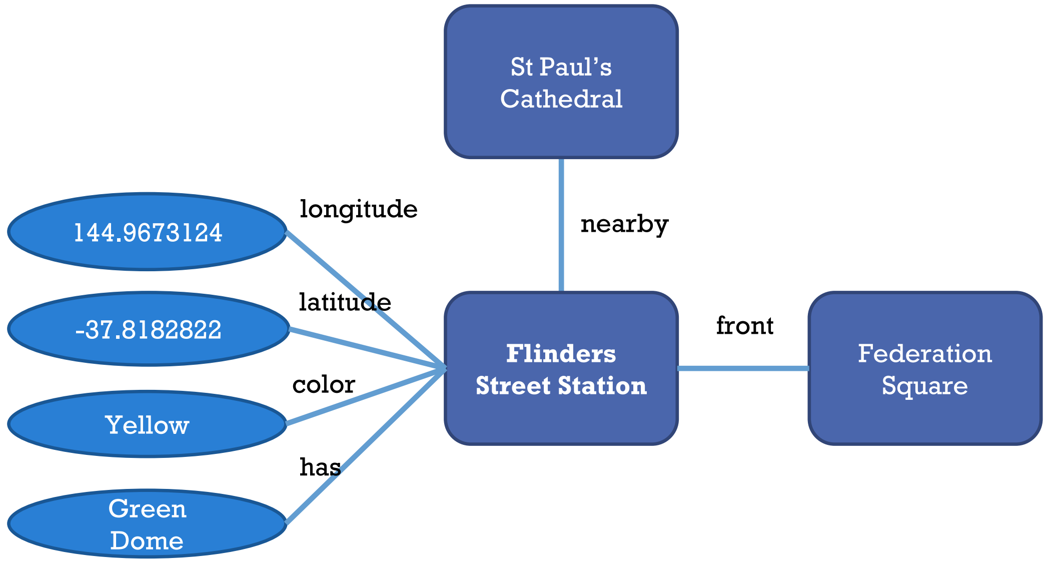
max planck institut informatik

Introduction

- A knowledge base is a large repository of facts, rules, and assumptions
- Scope: geographic knowledge base

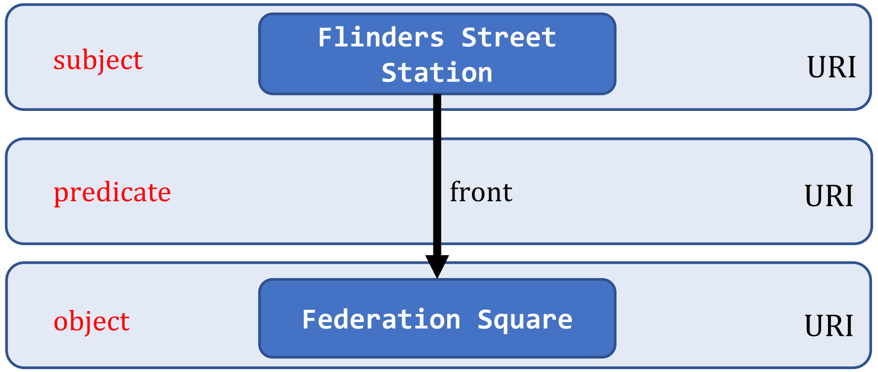
Introduction

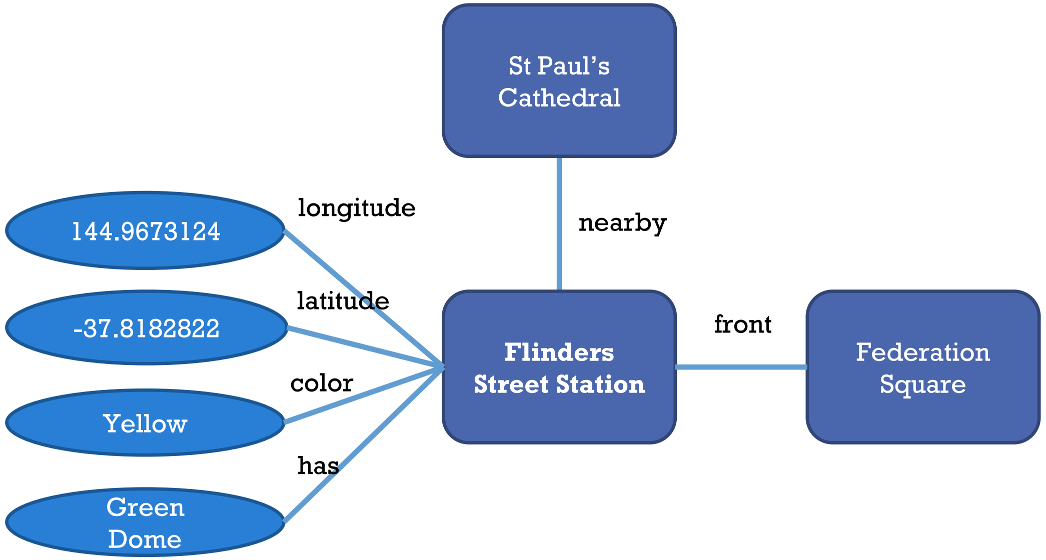
- A knowledge base is a large repository of facts, rules, and assumptions
- Scope: geographic knowledge base
- Representation:
  - Graph



Introduction

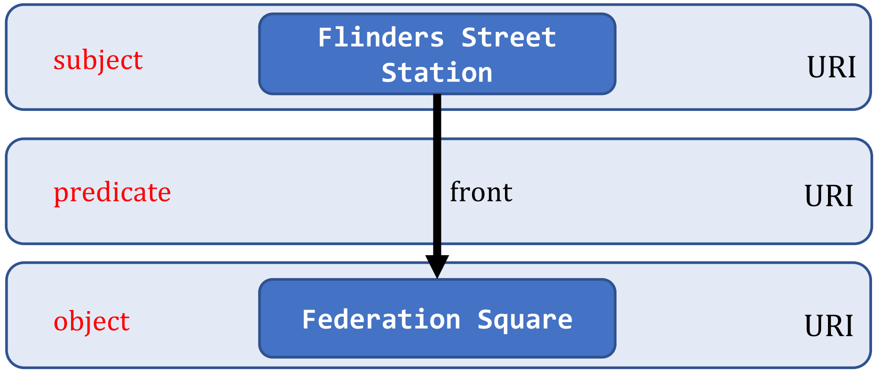
- A knowledge base is a large repository of facts, rules, and assumptions
- Scope: geographic knowledge base
- Representation:
  - Graph
  - Triple: `<Flinders Street Station, front, Federation Square>`

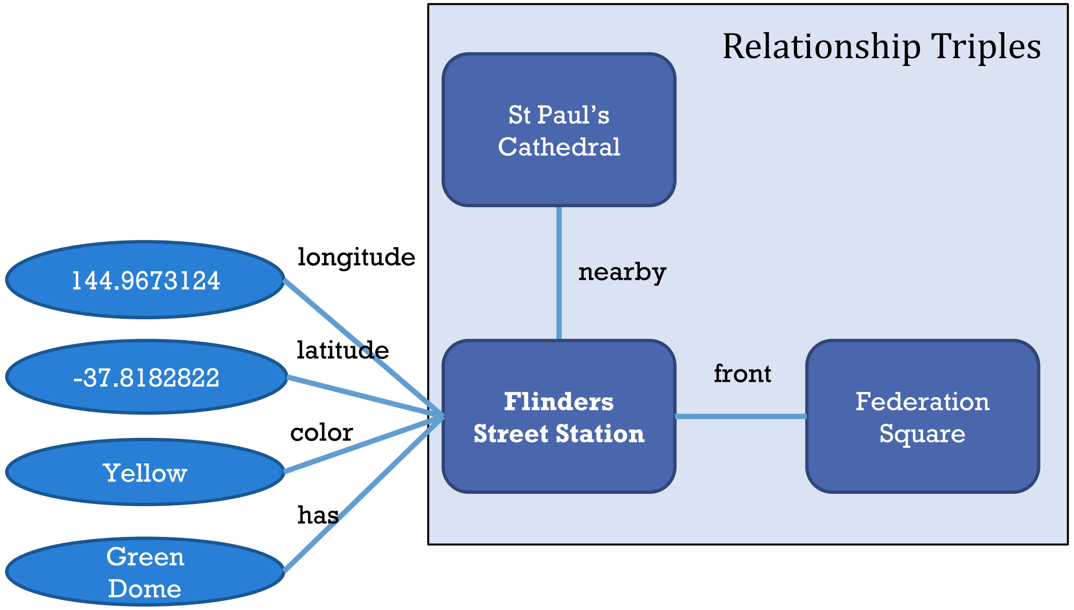




Introduction

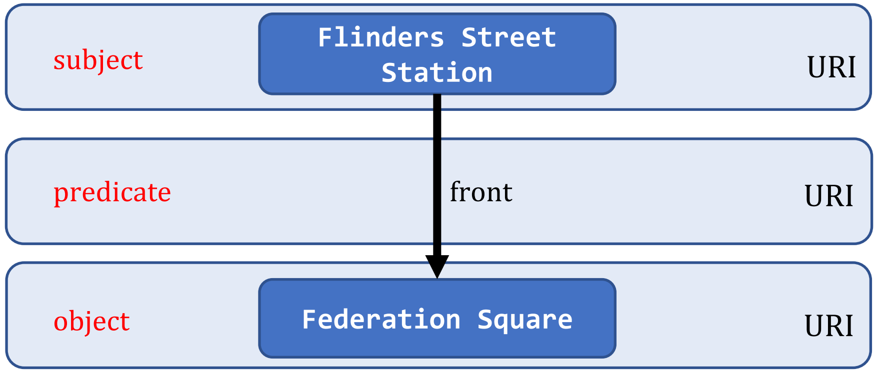
- A knowledge base is a large repository of facts, rules, and assumptions
- Scope: geographic knowledge base
- Representation:
  - Graph
  - Triple: `<Flinders Street Station, front, Federation Square>`

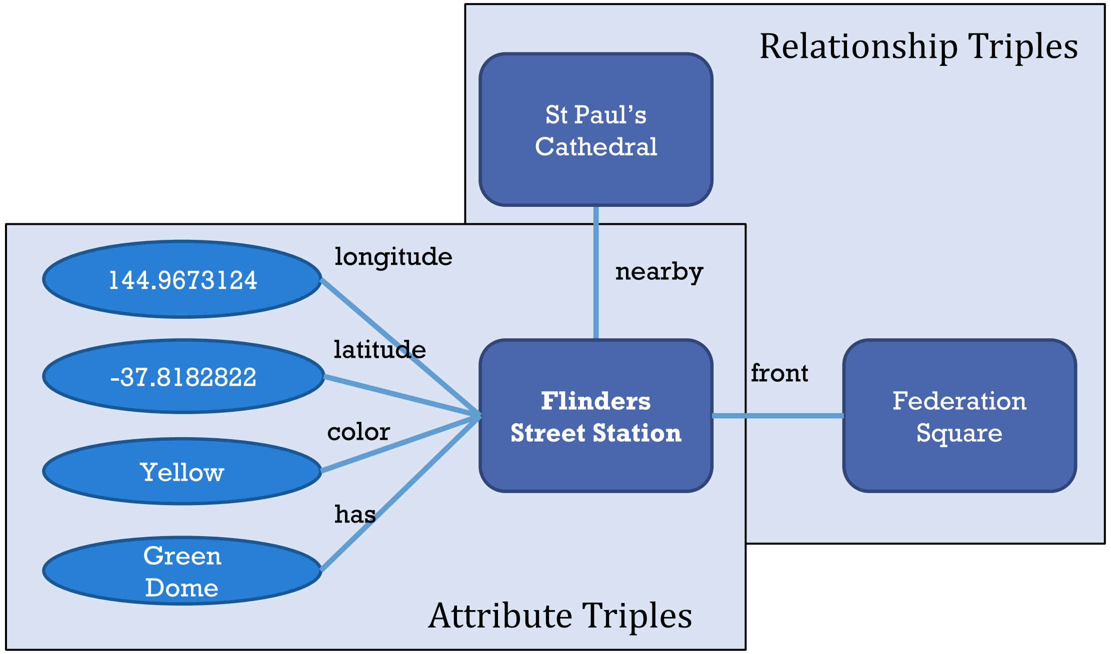




Introduction

- A knowledge base is a large repository of facts, rules, and assumptions
- Scope: geographic knowledge base
- Representation:
  - Graph
  - Triple: `<Flinders Street Station, front, Federation Square>`





Introduction

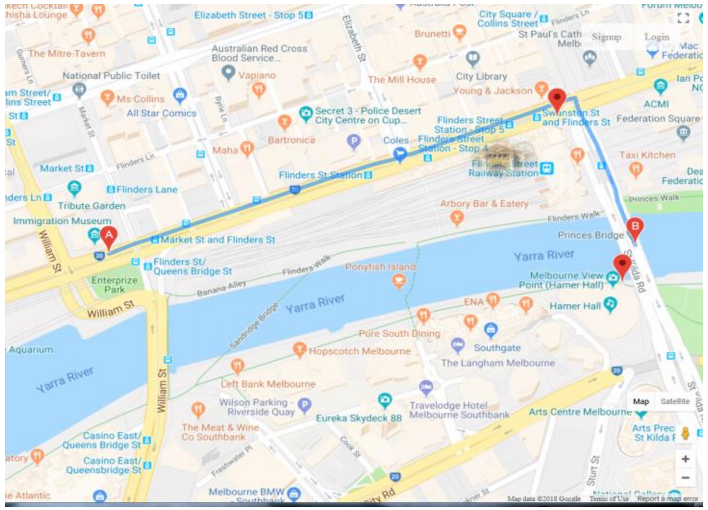
- Knowledge Base Usage
  - Location based information retrieval
    - *“Find an Italian restaurant near Melbourne University!”*

Introduction

- Knowledge Base Usage
  - Location based information retrieval
    - *"Find an Italian restaurant near Melbourne University!"*
  - Location based question answering
    - *"What is the tallest building in Melbourne?"*

Introduction

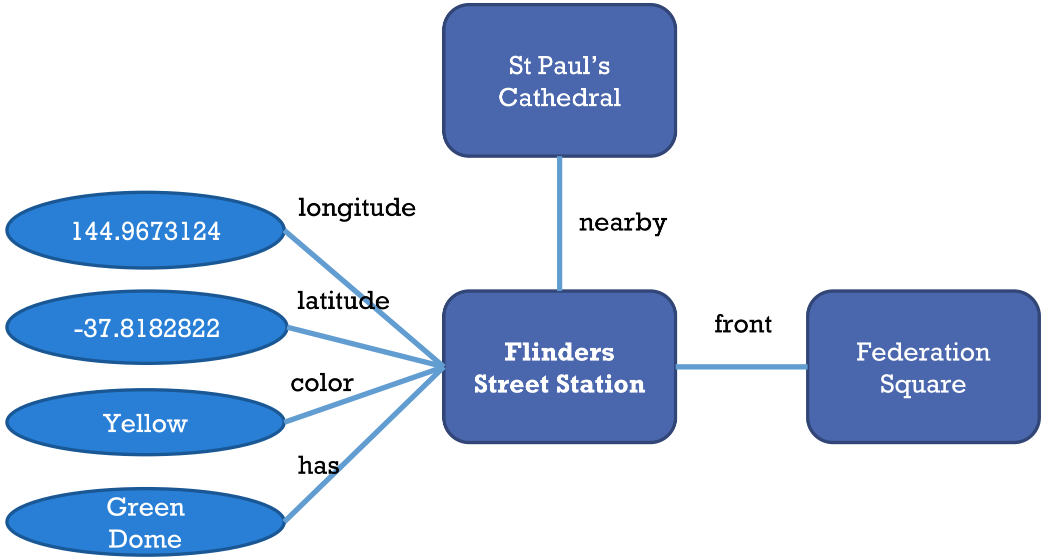
- Knowledge Base Usage
  - Location based information retrieval
    - *“Find an Italian restaurant near Melbourne University!”*
  - Location based question answering
    - *“What is the tallest building in Melbourne?”*
  - Navigation System

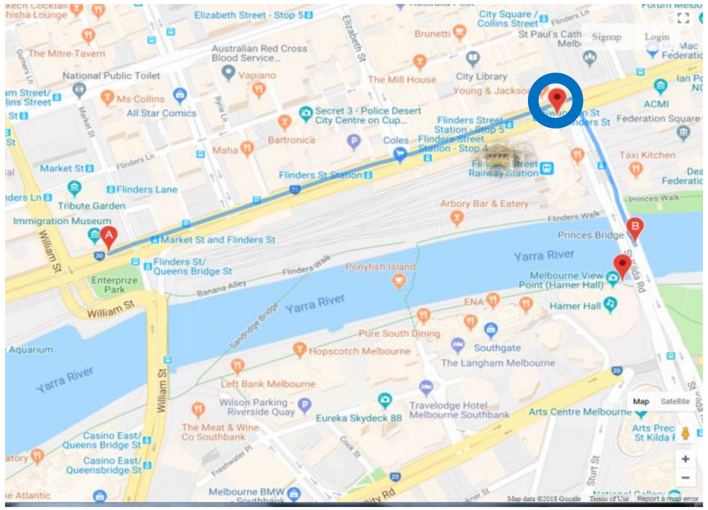


“Walk east on Flinders St/State Route 30 towards Market St; Turn right onto St Kilda Rd/Swanston St”

Introduction

- Knowledge Base Usage
  - Location based information retrieval
    - *"Find an Italian restaurant near Melbourne University!"*
  - Location based question answering
    - *"What is the tallest building in Melbourne?"*
  - Navigation System

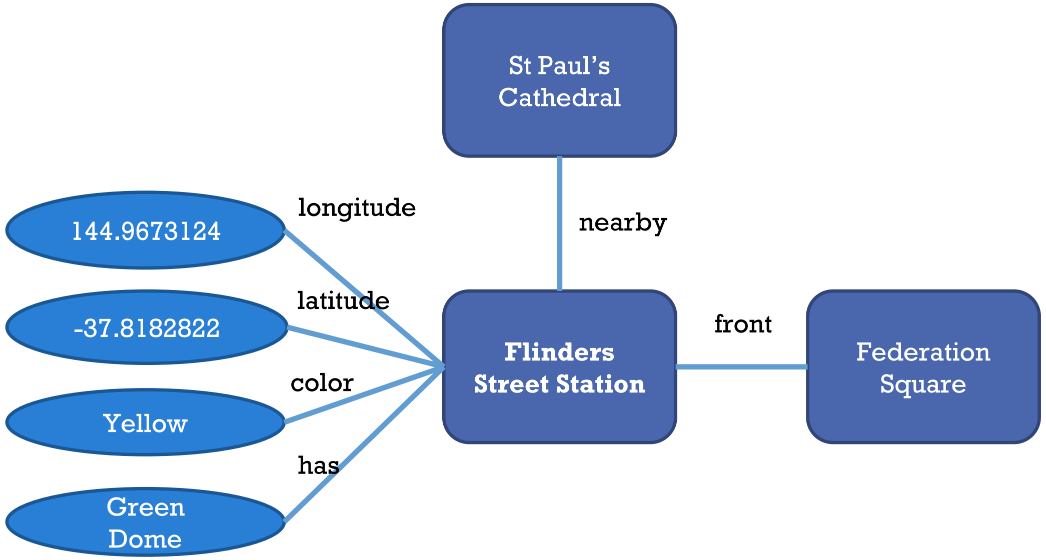


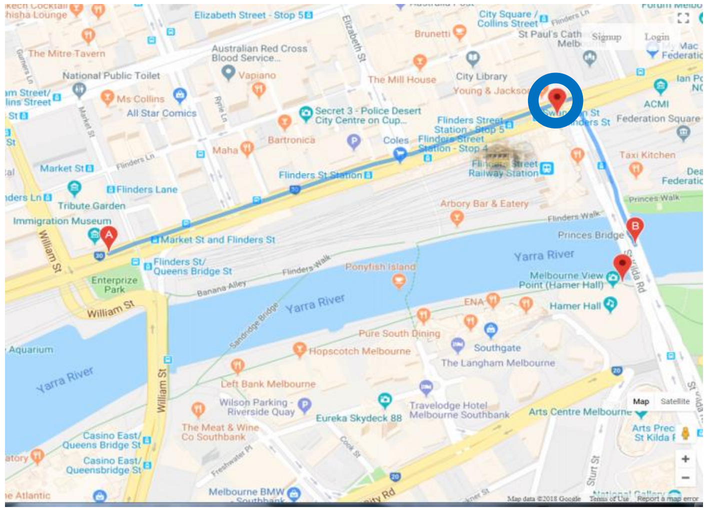


"Walk east on Flinders St/State Route 30 towards Market St; Turn right onto St Kilda Rd/Swanston St"

Introduction

- Knowledge Base Usage
    - Location based information retrieval
        - *“Find an Italian restaurant near Melbourne University!”*
    - Location based question answering
        - *“What is the tallest building in Melbourne?”*
    - Navigation System





“Walk east on Flinders St/State Route 30 towards Market St; Turn right onto St Kilda Rd/Swanston St”

vs.

“Walk east on Flinders St/State Route 30 towards Market St; Turn right onto St Kilda Rd/Swanston St after **Flinders Street Station, a yellow building with a green dome.”**

Motivation

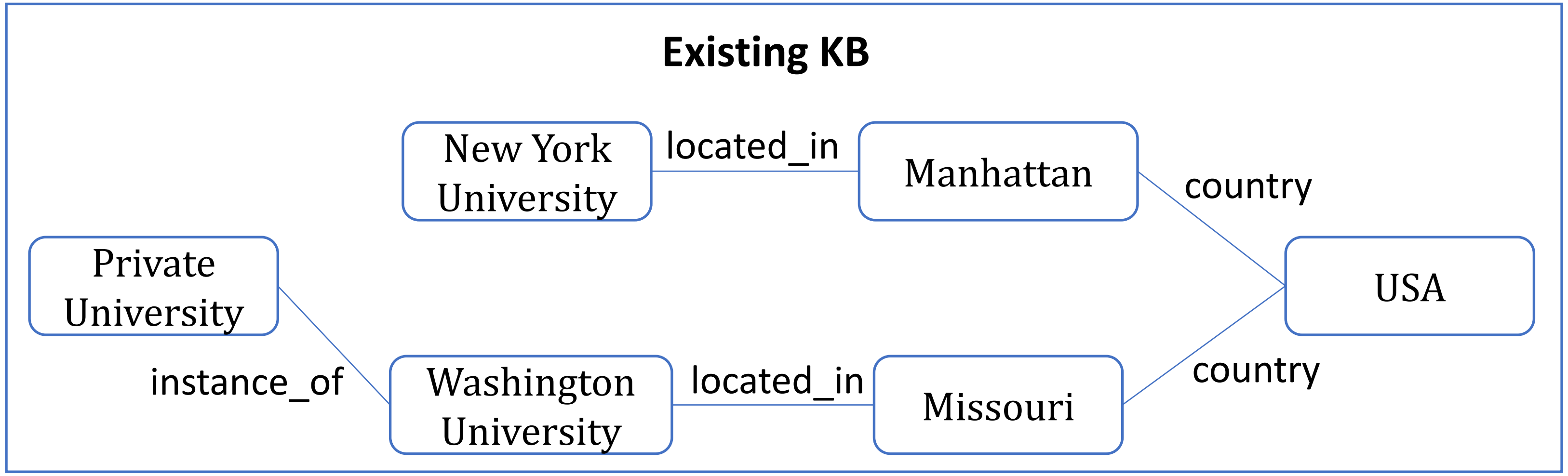
- Geographic knowledge base construction
  - Entity description generation (ACL 2018[1], AAAI 2020[2])
  - Knowledge base alignment (AAAI 2019[3])
  - **Knowledge base completion (ACL 2019[4])**

[1]Bayu Distiawan Trisedya, Jianzhong Qi, Rui Zhang, Wei Wang. 2018. GTR-LSTM: A Triple Encoder for Sentence Generation from RDF Data. *Proceedings of Annual Meeting of the Association for Computational Linguistics*.
[2]Bayu Distiawan Trisedya, Jianzhong Qi, Rui Zhang. 2020. Sentence Generation for Entity Description with Content-plan Attention, *Proceedings of AAAI Conference on Artificial Intelligence.*
[3]Bayu Distiawan Trisedya, Jianzhong Qi, Rui Zhang. 2019. Entity Alignment between Knowledge Graphs Using Attribute Embeddings, *Proceedings of AAAI Conference on Artificial Intelligence.*
[4]Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, Rui Zhang. 2019. Neural Relation Extraction for Knowledge Base Enrichment, *Proceedings of Annual Meeting of the Association for Computational Linguistics*.

*** Thesis: https://minerva-access.unimelb.edu.au/handle/11343/258706*

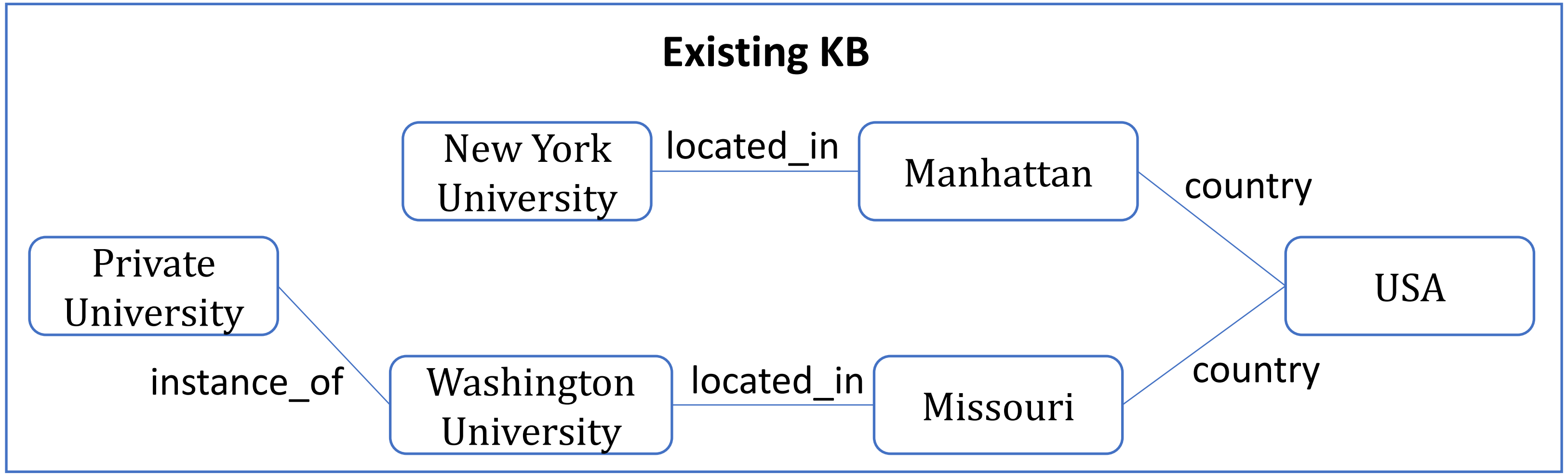
Motivation

- Knowledge base completion



Motivation

- Knowledge base completion



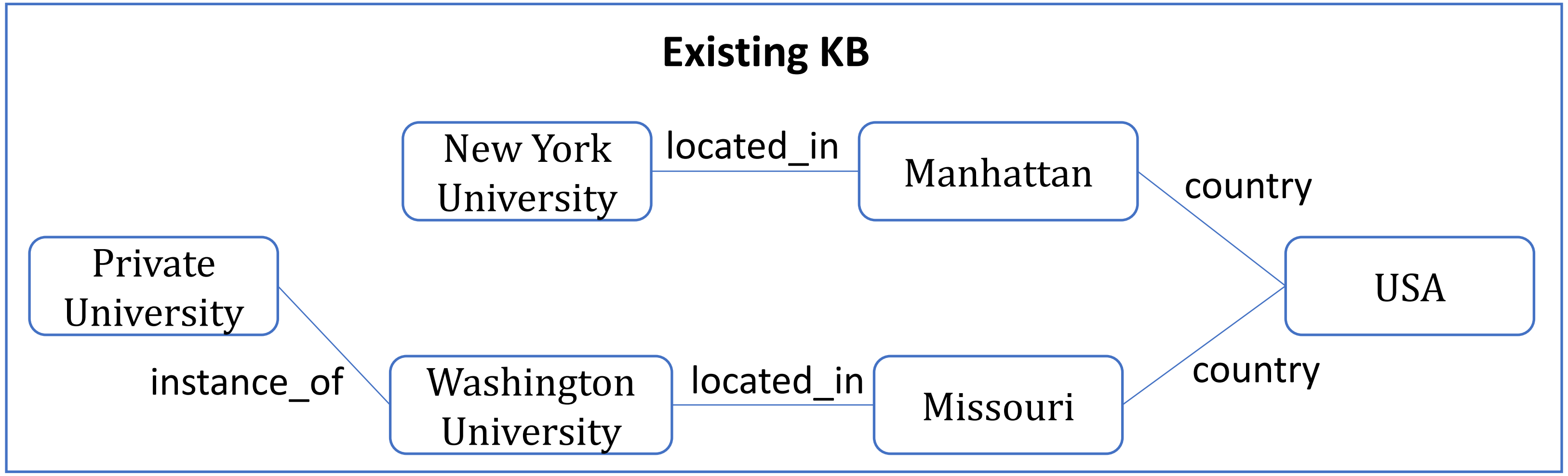
- Methods: Embedding-based knowledge base completion methods
  - Representational learning (a.k.a., entity embeddings)
  - We use these methods to verify the extracted facts

Motivation

- Evidence-based **Knowledge base completion**
  - Extract entities and their relationships from sentences to enrich an existing KB

Motivation

- Evidence-based **Knowledge base completion**
  - Extract entities and their relationships from sentences to enrich an existing KB
- Example:



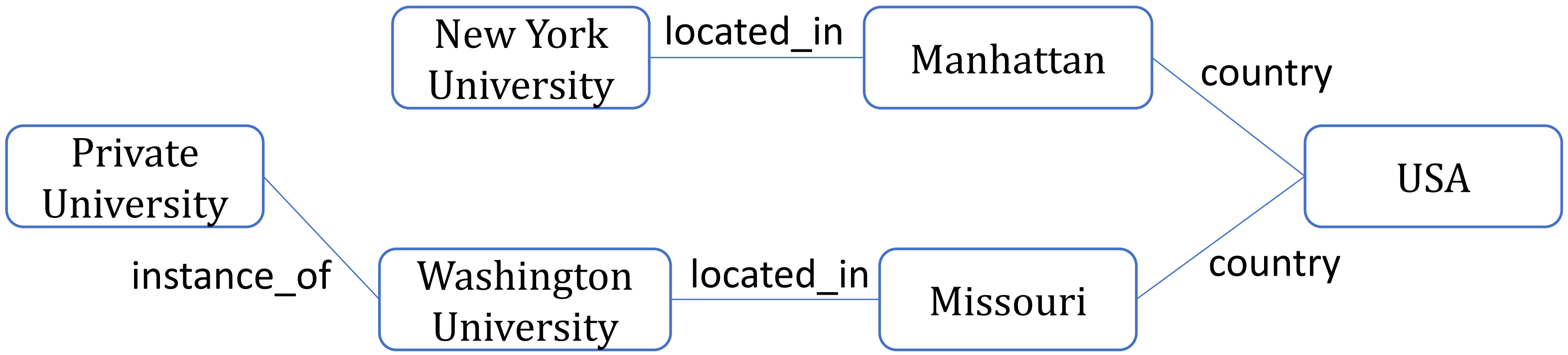
Relation Extraction

Sentence : New York University is a private university in Manhattan.

Motivation

- Evidence-based **Knowledge base completion**
  - Extract entities and their relationships from sentences to enrich an existing KB
- Example:

Existing KB



Relation Extraction

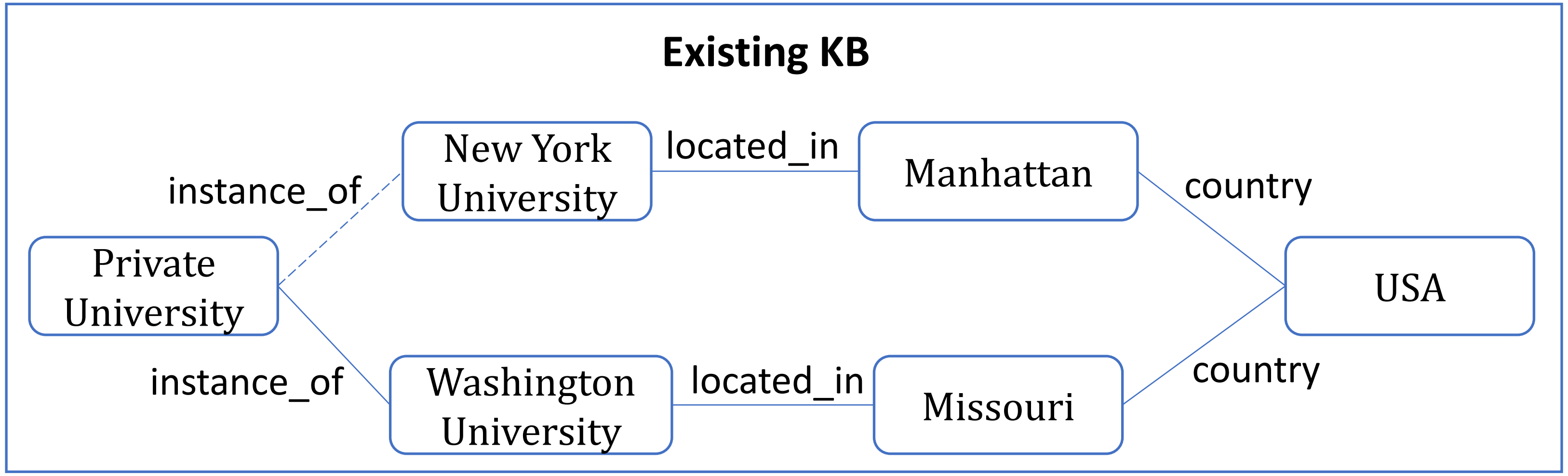
Sentence : New York University is a private university in Manhattan.

Output : `<New York University, instance_of, Private University>`

`<New York University, located_in, Manhattan>`

Motivation

- Evidence-based **Knowledge base completion**
  - Extract entities and their relationships from sentences to enrich an existing KB
- Example:



Relation Extraction

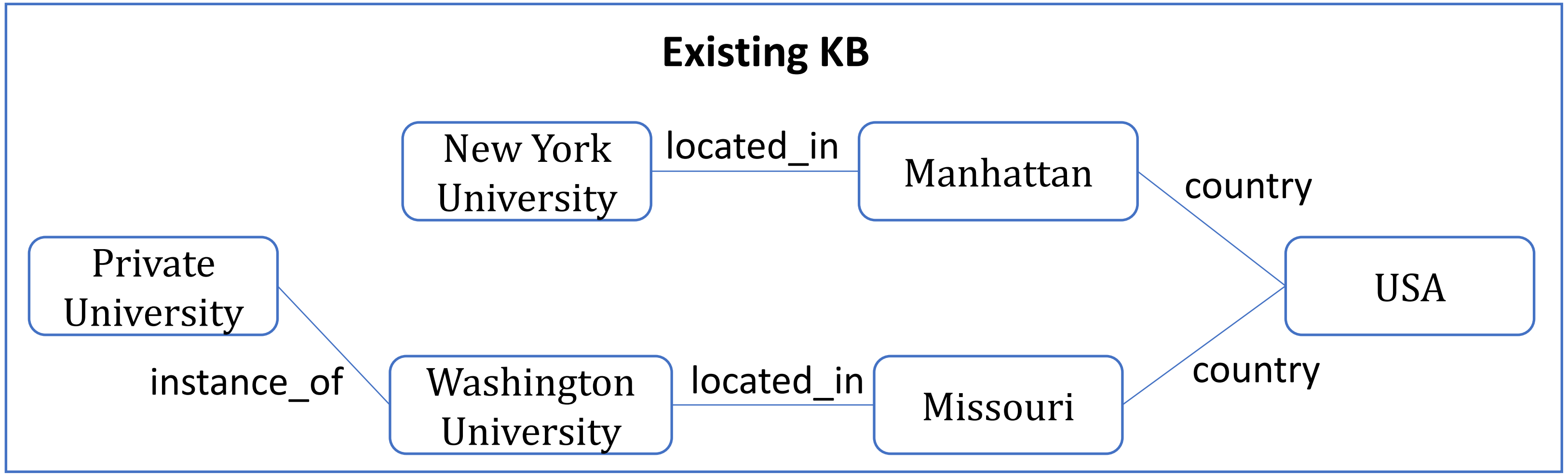
Sentence : New York University is a private university in Manhattan.

Output : **`<New York University, instance_of, Private University>`**

`<New York University, located_in, Manhattan>`

Related Work: Open Information Extraction

- Generate triples where the head and tail entities and the predicate stay in their surface forms
  - ReVerb (Fader et al., 2011), OLLIE (Mausam et al., 2012), ClausIE (Del Corro & Gemulla, 2013), MinIE (Gashteovski et al., 2017), RNN-OIE (Stanovsky et al., 2018), Neural-OIE (Cui et al., 2018)



Open IE

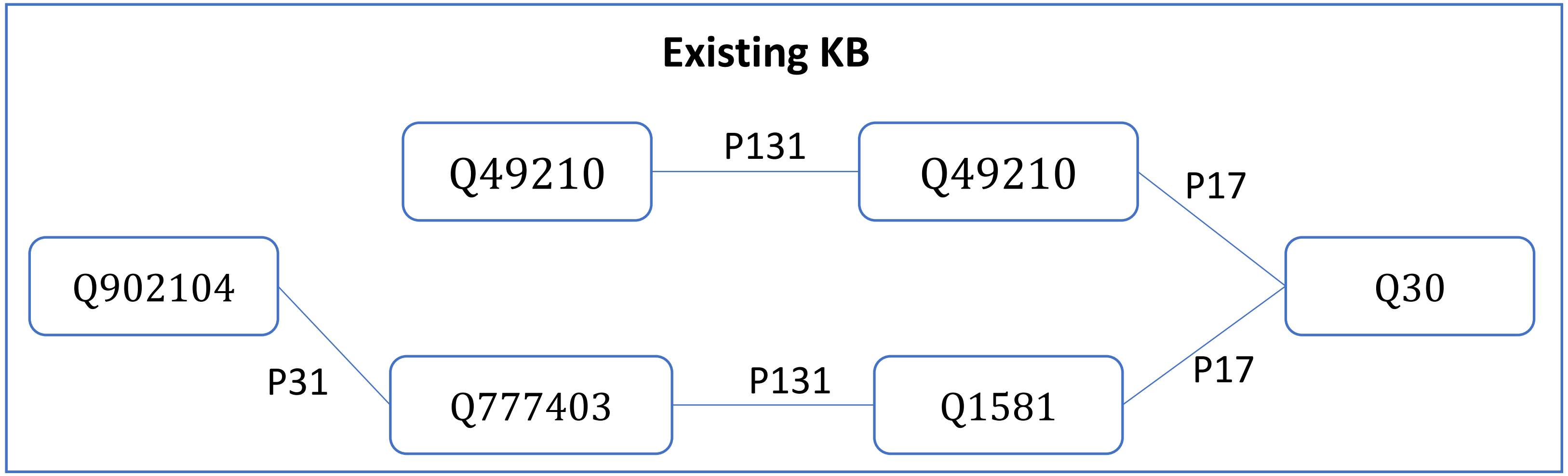
Sentence : New York University is a private university in Manhattan.

Output : **`<New York University, is, Private University>`**

`<New York University, is private university in, Manhattan>`

Related Work: Open Information Extraction

- Generate triples where the head and tail entities and the predicate stay in their surface forms
  - ReVerb (Fader et al., 2011), OLLIE (Mausam et al., 2012), ClausIE (Del Corro & Gemulla, 2013), MinIE (Gashteovski et al., 2017), RNN-OIE (Stanovsky et al., 2018), Neural-OIE (Cui et al., 2018)



Open IE

Sentence : New York University is a private university in Manhattan.

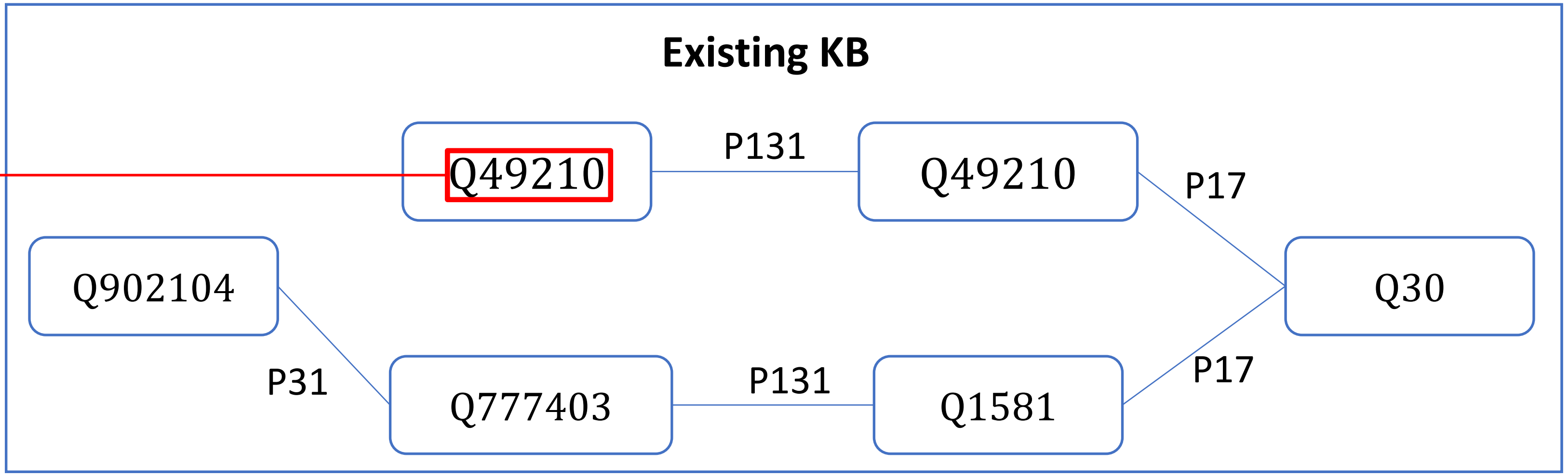
Output : **`<New York University, is, Private University>`**

`<New York University, is private university in, Manhattan>`

Related Work: Open Information Extraction

- Generate triples where the head and tail entities and the predicate stay in their surface forms
  - ReVerb (Fader et al., 2011), OLLIE (Mausam et al., 2012), ClausIE (Del Corro & Gemulla, 2013), MinIE (Gashteovski et al., 2017), RNN-OIE (Stanovsky et al., 2018), Neural-OIE (Cui et al., 2018)

Canonicalize the surface form using Named Entity Disambiguation (e.g., AIDA (Hoffart et al., 2011) and NeuralEL (Kolitsas et al., 2018).)



Open IE

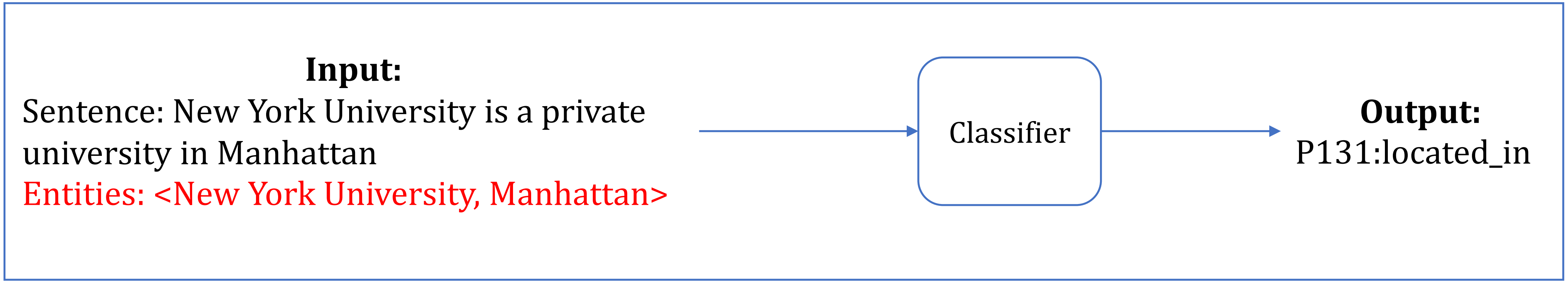
Sentence : New York University is a private university in Manhattan.

Output : **<New York University, is, Private University>**

<New York University, is private university in, Manhattan>

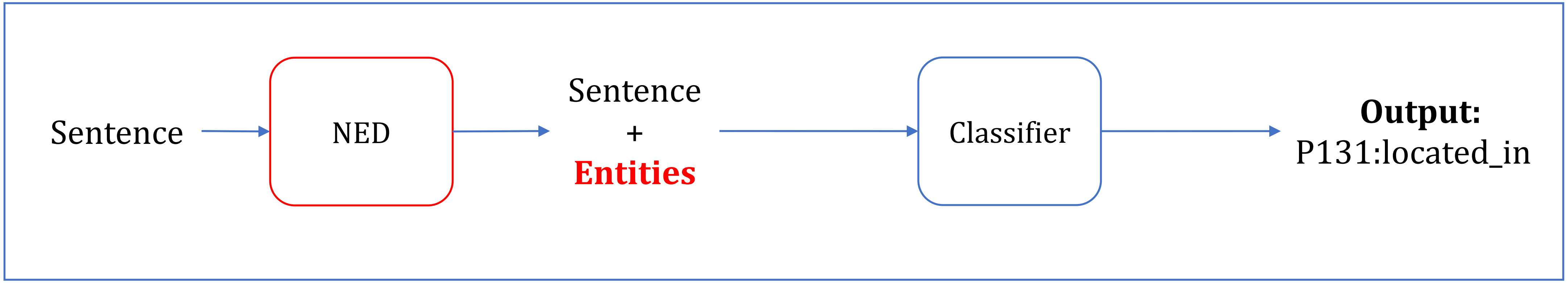
Related Work: Entity-aware Relation Extraction

- Learn extraction patterns from seed facts collected using distant supervision, then use statistical inference (e.g., classifier) to predict the relationship between two known entities.
  - **Traditional machine learning classifier** (Mintz et al. 2009; Hoffmann et al., 2010; Riedel et al., 2010, 2013; Surdeanu et al., 2012),
  - **CNN-based classifier** (Nguyen & Grishman 2015, Zeng et al. 2015, Lin et al. 2016),
  - **Attention-based classifier** (Zhou et al., 2018; Ji et al., 2017; Miwa and Bansal, 2016; Sorokin and Gurevych, 2017)



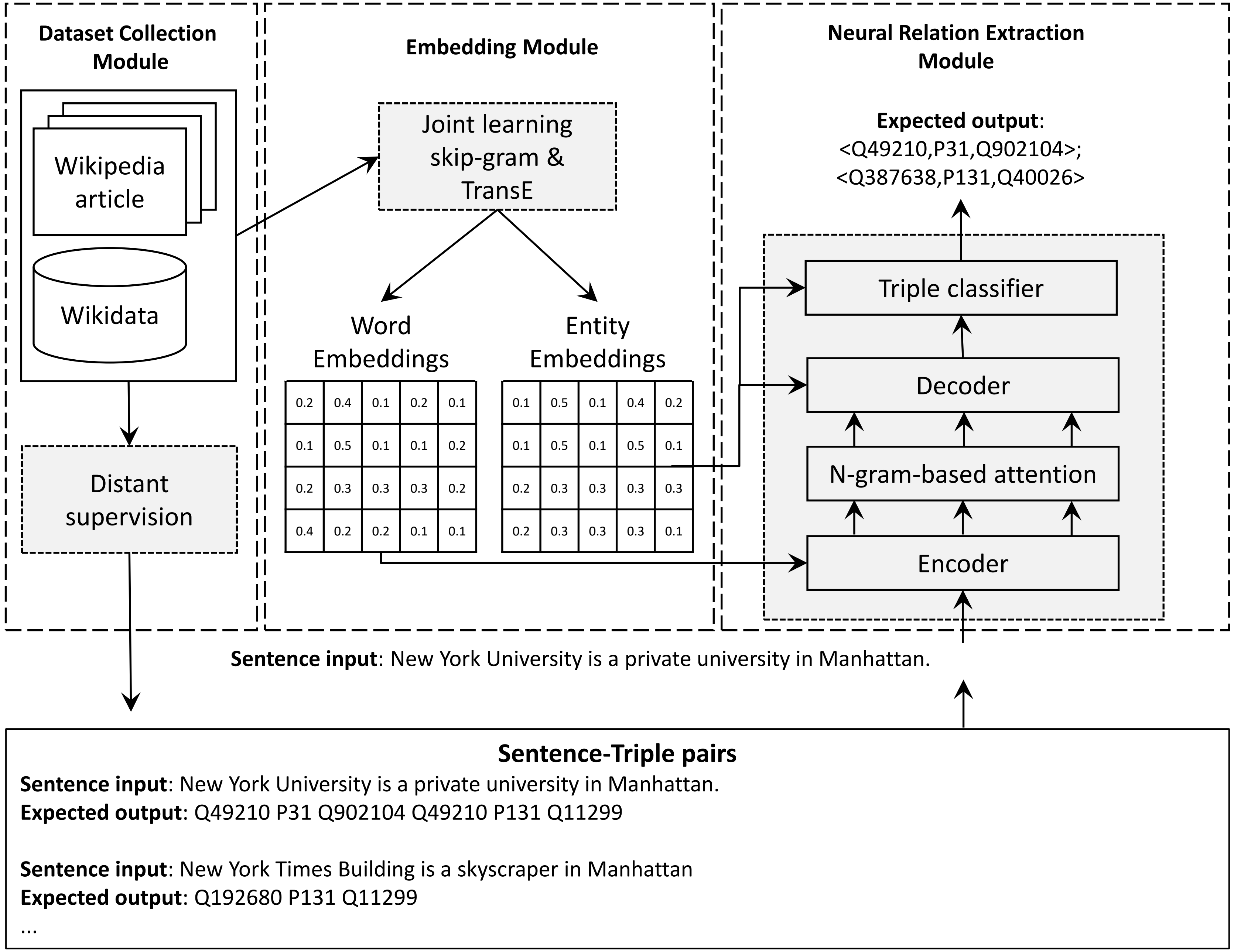
Related Work: Entity-aware Relation Extraction

- Learn extraction patterns from seed facts collected using distant supervision, then use statistical inference (e.g., classifier) to predict the relationship between two known entities.
  - **Traditional machine learning classifier** (Mintz et al. 2009; Hoffmann et al., 2010; Riedel et al., 2010, 2013; Surdeanu et al., 2012),
  - **CNN-based classifier** (Nguyen & Grishman 2015, Zeng et al. 2015, Lin et al. 2016),
  - **Attention-based classifier** (Zhou et al., 2018; Ji et al., 2017; Miwa and Bansal, 2016; Sorokin and Gurevych, 2017)



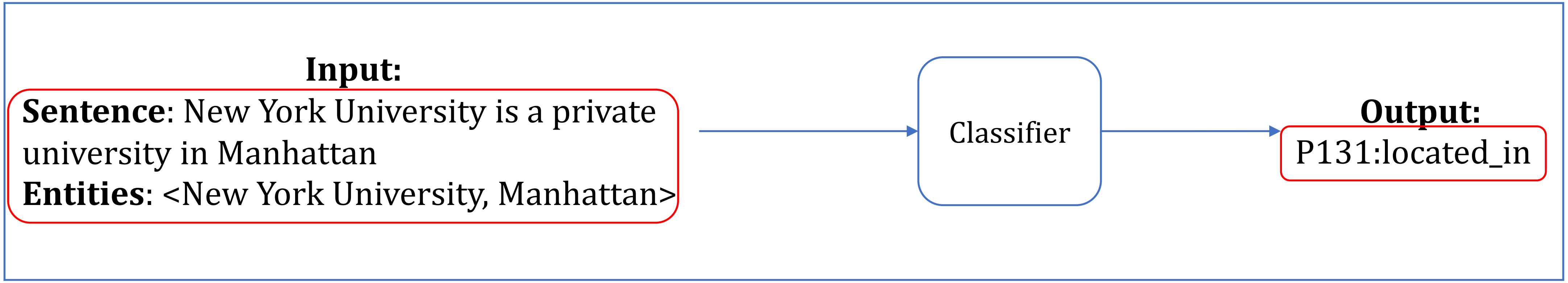
Proposed Model

- Existing methods rely on NED to map triples (both entity and relationship) into the existing KB
- We propose an end-to-end model for extracting and canonicalizing triples to enrich a KB.
  - Reduce error propagation between relation extraction and NED


Dataset Collection Module
Wikipedia article
Wikidata
Distant supervision
Embedding Module
Joint learning skip-gram & TransE
Word Embeddings
0.2 0.4 0.1 0.2 0.1
0.1 0.5 0.1 0.1 0.2
0.2 0.3 0.3 0.3 0.2
0.4 0.2 0.2 0.1 0.1
Entity Embeddings
0.1 0.5 0.1 0.4 0.2
0.1 0.5 0.1 0.5 0.1
0.2 0.3 0.3 0.3 0.3
0.2 0.3 0.3 0.3 0.1
Neural Relation Extraction Module
Expected output:
<Q49210,P31,Q902104>;
<Q387638,P131,Q40026>
Triple classifier
Decoder
N-gram-based attention
Encoder
Sentence input: New York University is a private university in Manhattan.
Sentence-Triple pairs
Sentence input: New York University is a private university in Manhattan.
Expected output: Q49210 P31 Q902104 Q49210 P131 Q11299
Sentence input: New York Times Building is a skyscraper in Manhattan
Expected output: Q192680 P131 Q11299
...

Preliminary: Distant Supervision RE

- Machine learning models require extensive training data
- Mintz et al. (2009) propose distant supervision technique to **automatically** obtain large training data for relation extraction



Preliminary: Distant Supervision RE

Text corpus: Wikipedia

…

Barack Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago. In 1988, he enrolled in Harvard Law School, where he was the first black person to be president of the *Harvard Law Review*. …

…

Knowledge Base: Wikidata

```
...
<Barack Obama, place_of_birth, Honolulu>
<Keanu Reeves, place_of_birth, Beirut>
...
```

Preliminary: Distant Supervision RE

Text corpus: Wikipedia

…

Barack Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago. In 1988, he enrolled in Harvard Law School, where he was the first black person to be president of the *Harvard Law Review*. …

…

Knowledge Base: Wikidata

...
<Barack Obama, place_of_birth, Honolulu>
<Keanu Reeves, place_of_birth, Beirut>
...

Preliminary: Distant Supervision RE

Text corpus: Wikipedia

…

Barack Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago. In 1988, he enrolled in Harvard Law School, where he was the first black person to be president of the *Harvard Law Review*. …

…

Knowledge Base: Wikidata

```
...
<Barack Obama, place_of_birth, Honolulu>
<Keanu Reeves, place_of_birth, Beirut>
...
```

```
Dataset:
{
  {
    "sentence": "Barack Obama was born in Honolulu, Hawaii.",
    "entity": [{"Barack Obama", "Honolulu"}],
    "output": "place_of_birth"
  }
}
```

Embeddings – Representational Learning

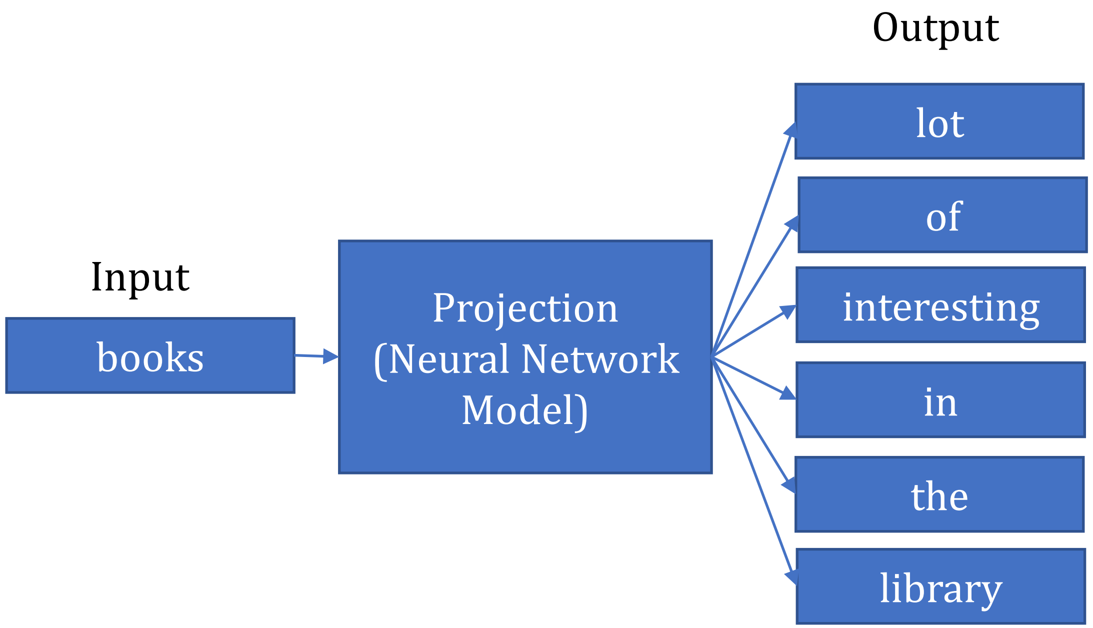
- A word/entity is represented as a vector.
  - Similarity is computed using cosine.
- Traditional word embedding methods:
  - Bag of words (one hot encoding)
    - For example, if we have a vocabulary of 10000 words, and “book” is the 2nd word in the dictionary, it would be represented by: [0 1 0 . . . . 0 0]
  - Document representation (e.g., TF-IDF)
  - Lack of context representation

Embeddings – Representational Learning

- Word embeddings
  - Dense vector of fixed number of dimensions (e.g., 300).
    - For example, representation of “book”: [0.5, -0.27, 0.75, 0.8 . . . 0.02, 0.1]
  - Methods
    - **Word2Vec** (2013)
    - GloVe (2014)
    - ELMo (2018)

Embeddings – Representational Learning

- Word2Vec
  - Skip gram
  - Example: "There are a lot of interesting **books** in the library."



Embeddings – Representational Learning

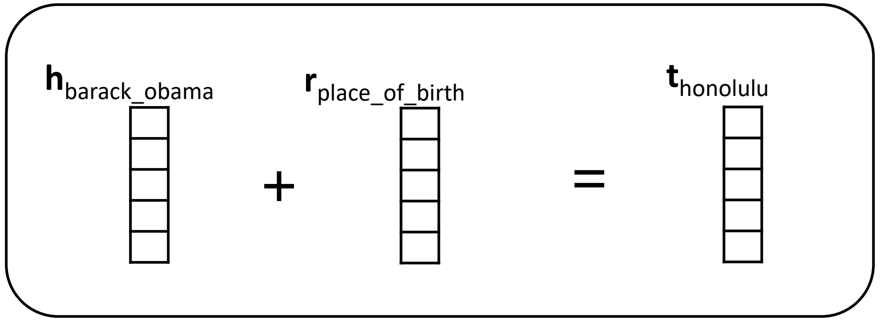
- Entity embeddings
  - Assigns a vector representation for each entity in a KG
- Models
  - **TransE** (2013)
  - KG2E (2015)
  - DistMult (2015)
  - Graph Convolutional Networks (2017)
  - Graph Attention Networks (2018)

Embeddings – Representational Learning

- TransE
  - For each valid triple in a KB, it assume:

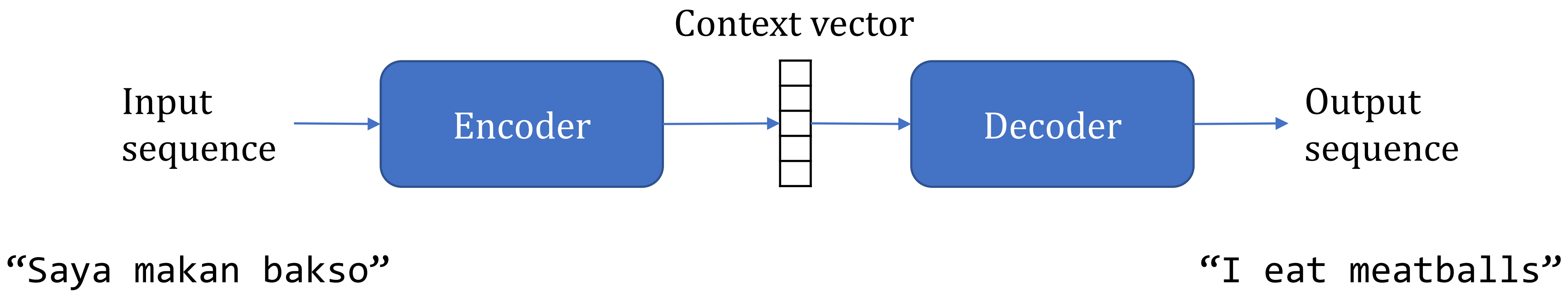
Knowledge Base: Wikidata

```
...
<Barack Obama, place_of_birth, Honolulu>
<Keanu Reeves, place_of_birth, Beirut>
...
```

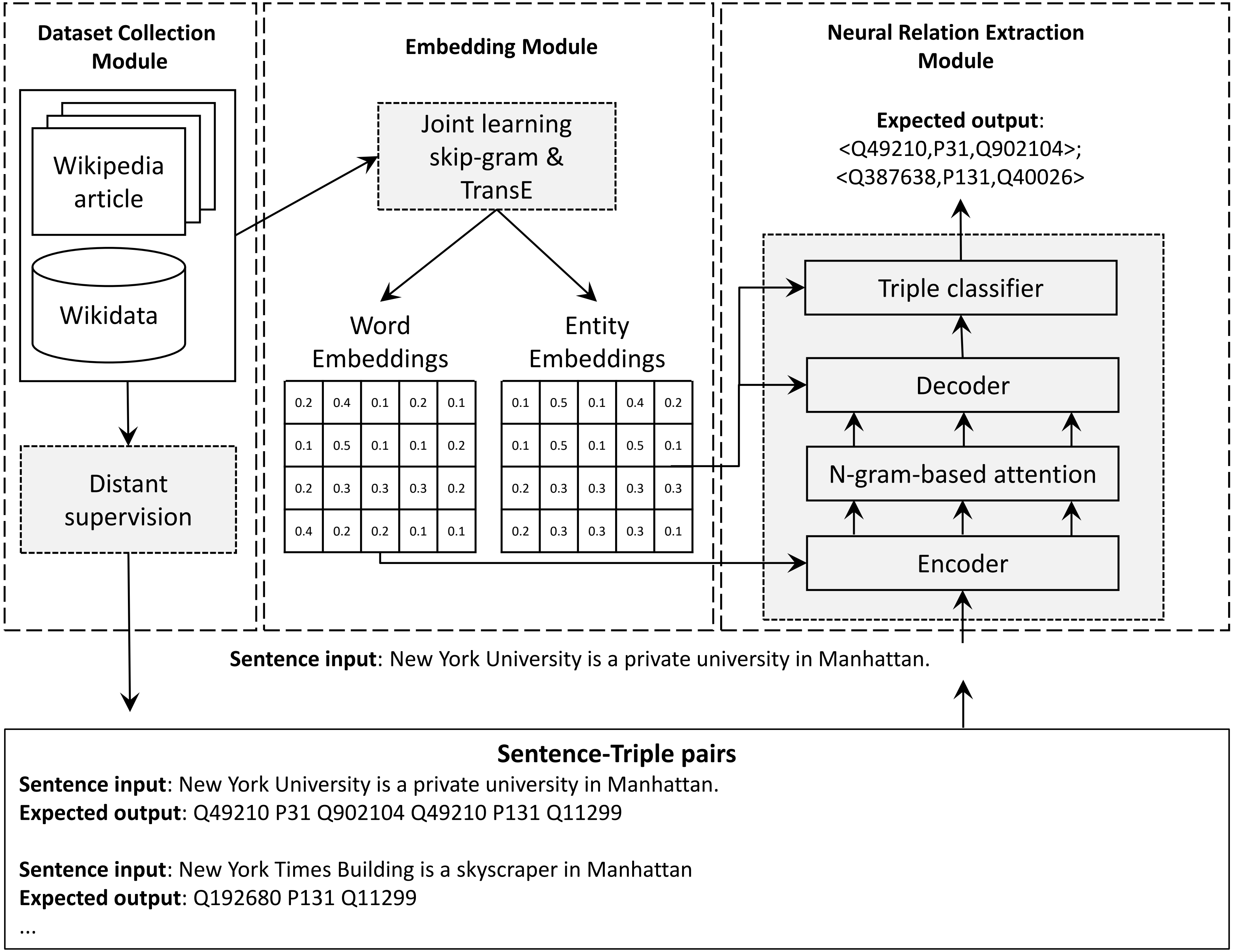


Encoder-Decoder Framework

- Sequence-to-sequence learning framework
  - Early models: Sutskever et al. (2014), Cho et al. (2014)



  - The encoder/decoder is a sequence model
    - RNN, LSTM, GRU, Transformer
  - Successfully applied for neural machine translation


Dataset Collection Module
Wikipedia article
Wikidata
Distant supervision
Embedding Module
Joint learning skip-gram & TransE
Word Embeddings
Entity Embeddings
0.2 0.4 0.1 0.2 0.1
0.1 0.5 0.1 0.1 0.2
0.2 0.3 0.3 0.3 0.2
0.4 0.2 0.2 0.1 0.1
0.1 0.5 0.1 0.4 0.2
0.1 0.5 0.1 0.5 0.1
0.2 0.3 0.3 0.3 0.3
0.2 0.3 0.3 0.3 0.1
Neural Relation Extraction Module
Expected output:
<Q49210,P31,Q902104>;
<Q387638,P131,Q40026>
Triple classifier
Decoder
N-gram-based attention
Encoder
Sentence input: New York University is a private university in Manhattan.
Sentence-Triple pairs
Sentence input: New York University is a private university in Manhattan.
Expected output: Q49210 P31 Q902104 Q49210 P131 Q11299
Sentence input: New York Times Building is a skyscraper in Manhattan
Expected output: Q192680 P131 Q11299
...

Proposed Model: Dataset Collection

- Goals: create **large volume** of fully labeled **high-quality** training data in the form of sentence-triple pairs.
- **Large volume**: extracting sentences that contain implicit entity names using co-reference resolution (Clark and Manning, 2016)

```
Barack Obama is an American attorney and politician … He was reelected
                  to the Illinois Senate …
```

↓

```
Barack Obama is an American attorney and politician … Barack Obama was
              reelected to the Illinois Senate …
```

Proposed Model: Dataset Collection

- Goals: create **large volume** of fully labeled **high-quality** training data in the form of sentence-triple pairs.
- **High-quality**: filtering sentences that do not express any relationships using paraphrase detection (PATTY (Nakashole et al., 2012), POLY (Grycner and Weikum, 2016), and PPDB (Ganitkevitch et al., 2013))

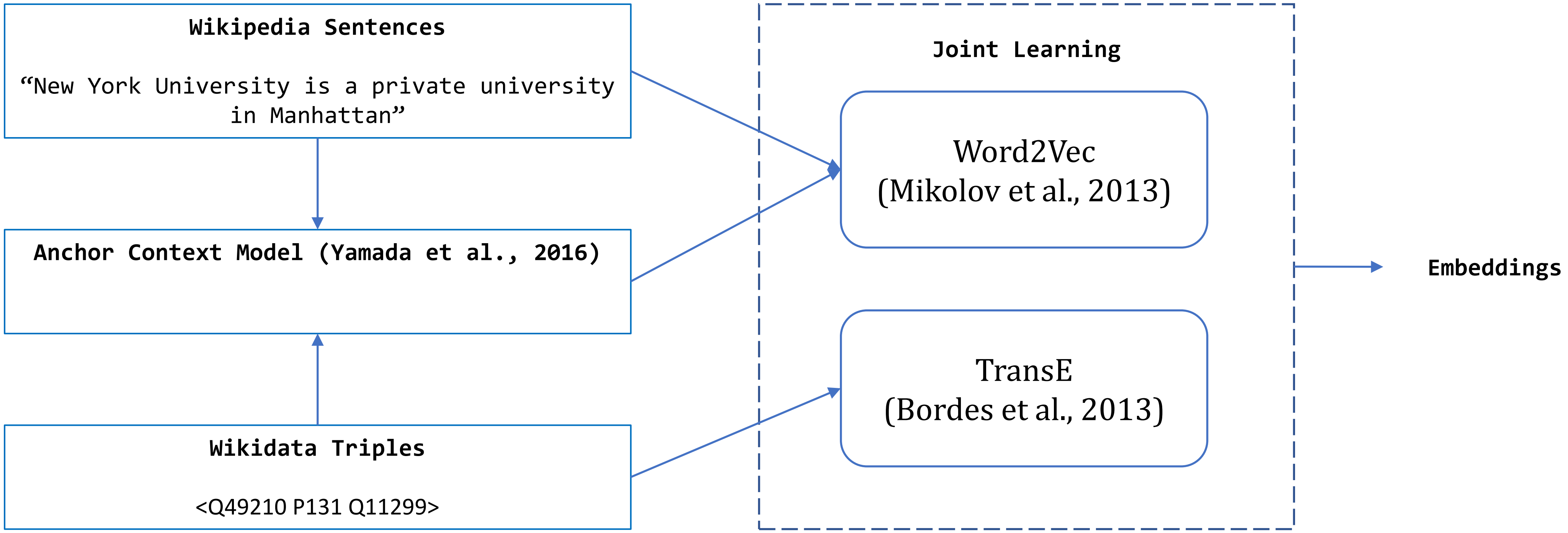
<Barack Obama, place_of_birth, Honolulu>

☑ Barack Obama was born in 1961 in Honolulu, Hawaii

☒ Barack Obama visited Honolulu in 2010

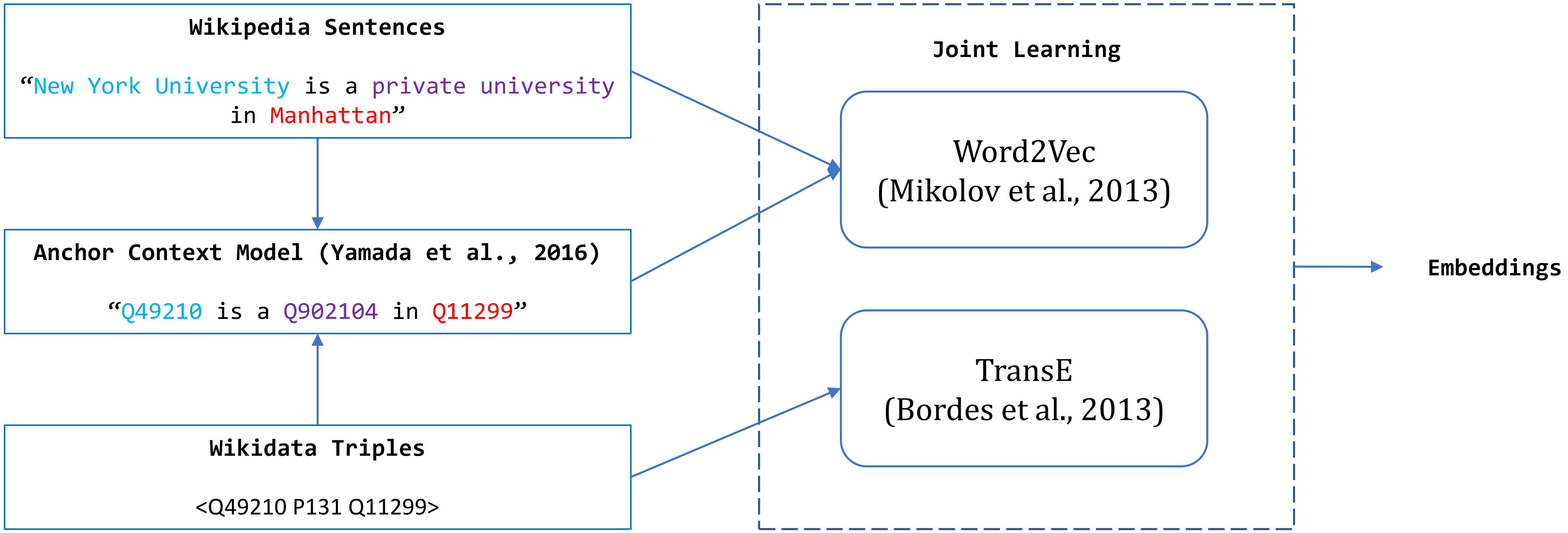
Proposed Model: Embedding Module

- Goals:
  - Compute pre-trained embeddings
  - Capture the similarity between words and entities for named entity disambiguation (Yamada et al., 2016)



Proposed Model: Embedding Module

- Goals:
  - Compute pre-trained embeddings
  - Capture the similarity between words and entities for named entity disambiguation (Yamada et al., 2016)



Proposed Model: Neural Relation Extraction Module

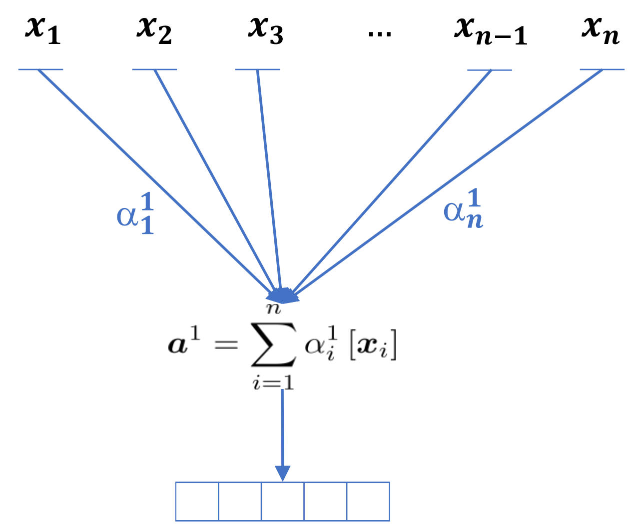
- Architecture: sequence to sequence model using encoder-decoder framework (Bahdanau et al., 2015)
  - To capture multi-words entity names, we propose **n-gram based attention model**
  - To generate high-quality triples, we use **triple classifier**

Proposed Model: Neural Relation Extraction Module

- **N-gram based attention model**
  - 82.9% of entity in the dataset have multi-words entity name
  - Expand the attention model to lookup all possible n-gram

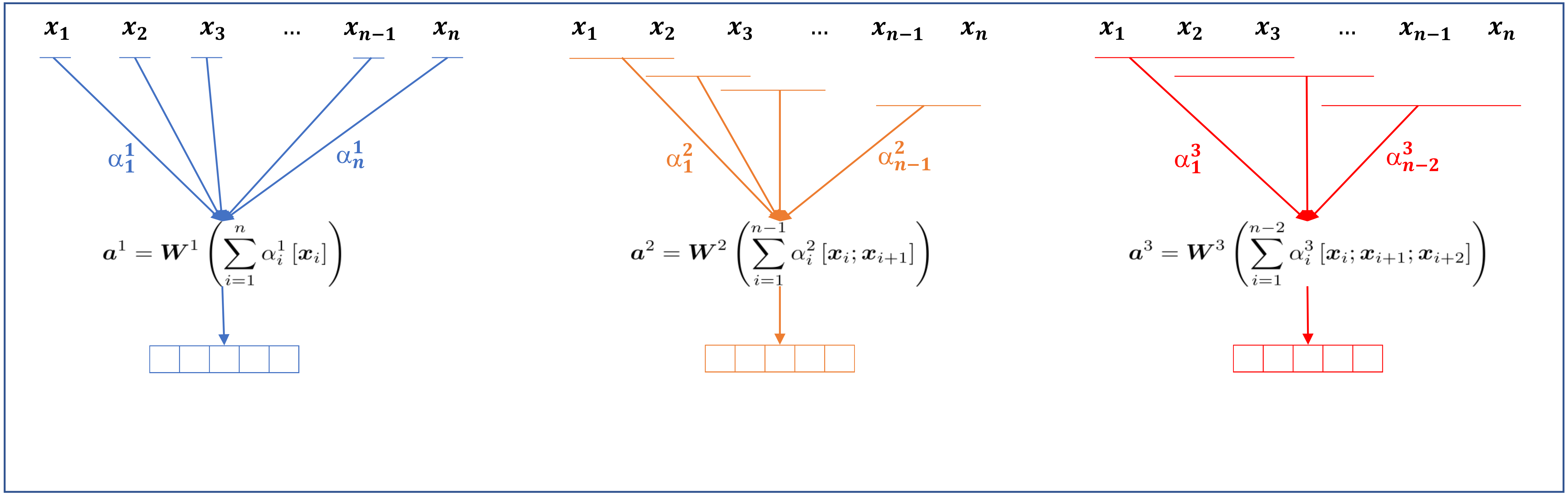
Proposed Model: Neural Relation Extraction Module

- **N-gram based attention model**
  - 82.9% of entity in the dataset have multi-words entity name
  - Expand the attention model to lookup all possible n-gram



Proposed Model: Neural Relation Extraction Module

- **N-gram based attention model**
  - 82.9% of entity in the dataset have multi-words entity name
  - Expand the attention model to lookup all possible n-gram



Proposed Model: Neural Relation Extraction Module

- **Triple classifier**
  - We train a binary classifier based on the plausibility score $(h + r - t)$, the score to compute the entity embeddings using TransE.
  - We create negative samples by corrupting the valid triples (i.e., replacing the head or tail entity by a random entity)
  - The triple classifier is effective to filter invalid triple such as
    - `<New York University, capital_of, Manhattan>`.

Experiments

- Dataset:
  - Data collected by the dataset collection module (**WIKI** dataset)
  - For stress testing, we also collect another test dataset outside Wikipedia.
    - We apply the same procedure to the user reviews of a travel website
    - Collect 1,000 sentence-triple pairs for test dataset (**GEO** dataset)

	#pairs	#triple	#entity	#relation type
All (WIKI)	255,654	330,005	279,888	158
Train+val	225,869	291,352	249,272	157
Test (WIKI)	29,785	38,653	38,690	109
Test (GEO)	1,000	1,095	124	11

Experiments

- Model:
  - CNN (the state-of-the-art supervised approach by Lin et al. (2016))
  - MiniE (the state-of-the-art unsupervised approach by Gashteovski et al. (2017))
  - ClausIE (the predecessor of Minnie by Corro and Gemulla (2013))
  - Single Attention model (Sequence to sequence model by Bahdanau et al., 2015)
  - Transformer model (Sequence to sequence model by Vaswani et al., 2017)
  - N-gram Attention model (Proposed model)

- For CNN, MiniE, and ClausIE, we use the following tools to canonicalize entity name and relationship:
  - NED tools AIDA (Hoffart et al., 2011) and NeuralEL (Kolitsas et al., 2018), to map the entity name
  - Paraphrase detection (PATTY (Nakashole et al., 2012), POLY (Grycner and Weikum, 2016), and PPDB (Ganitkevitch et al., 2013)) to map the relationship

Experiments

Model		WIKI Precision	WIKI Recall	WIKI F1	GEO Precision	GEO Recall	GEO F1
Existing Models	MinIE (+AIDA)	0.3672	0.4856	0.4182	0.3574	0.3901	0.3730
	MinIE (+NeuralEL)	0.3511	0.3967	0.3725	0.3644	0.3811	0.3726
	ClausIE (+AIDA)	0.3617	0.4728	0.4099	0.3531	0.3951	0.3729
	ClausIE (+NeuralEL)	0.3445	0.3786	0.3607	0.3563	0.3791	0.3673
	CNN (+AIDA)	0.4035	0.3503	0.3750	0.3715	0.3165	0.3418
	CNN (+NeuralEL)	0.3689	0.3521	0.3603	0.3781	0.3005	0.3349
Encoder-Decoder Models	Single Attention	0.4591	0.3836	0.4180	0.4010	0.3912	0.3960
	Single Attention (+pre-trained)	0.4725	0.4053	0.4363	0.4314	0.4311	0.4312
	Single Attention (+triple classifier)	0.7378	0.5013	0.5970	0.6704	0.5301	0.5921
	Transformer	0.4628	0.3897	0.4231	0.4575	0.4620	0.4597
	Transformer (+pre-trained)	0.4748	0.4091	0.4395	0.4841	0.4831	0.4836
	Transformer (+triple classifier)	0.7307	0.4866	0.5842	0.7124	0.5761	0.6370
Proposed	N-gram Attention	0.7014	0.6432	0.6710	0.6029	0.6033	0.6031
	N-gram Attention (+pre-trained)	0.7157	0.6634	0.6886	0.6581	0.6631	0.6606
	N-gram Attention (+triple classifier)	**0.8471**	0.6762	**0.7521**	**0.7705**	0.6771	**0.7208**

Summary

- Neural relation extraction model for KB completion.
  - Dataset collection: distant supervision + co-reference resolution + paraphrase detection
  - Embedding module: Word2Vec + TransE + Anchor Context Model
  - Extraction module: Encoder-decoder + N-gram attention + triple classification
- Future work:
  - Our model handle syntactic similarity → Context-based similarity?
  - Knowledge base enrichment?