BOND:
BERT-Assisted **O**pen-Domain **N**amed Entity Recognition with **D**istant Supervision

Chen Liang*, Yue Yu*, Haoming Jiang*,
Siawpeng Er, Ruijia Wang, Tuo Zhao, Chao Zhang

Georgia Institute of Technology

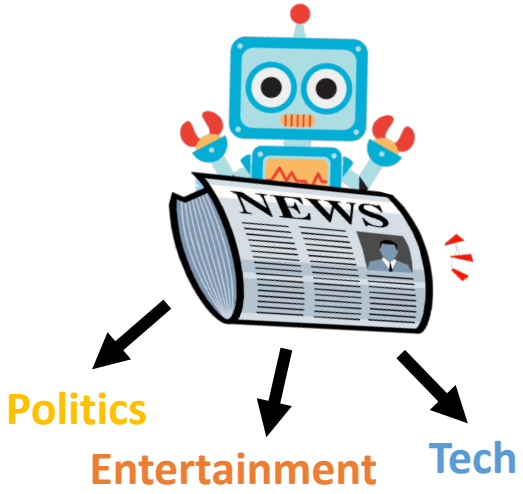
* Equal contributions

Named Entity Recognition (NER)

When Sebastian Thrun PERSON started at Google ORG in 2007 DATE , few people outside of the company took him seriously. “I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn’t worth talking to,” said Thrun PERSON , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE .

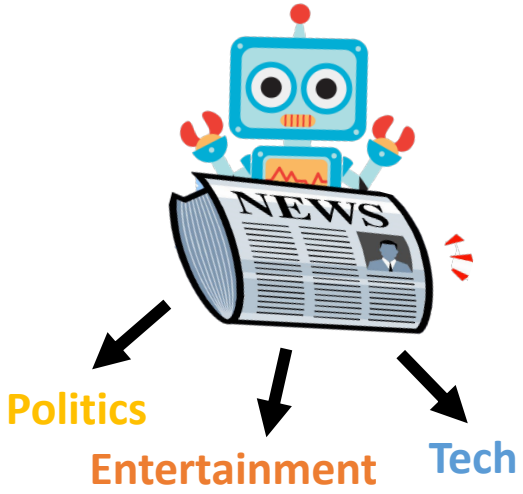
Named Entity Recognition (NER)

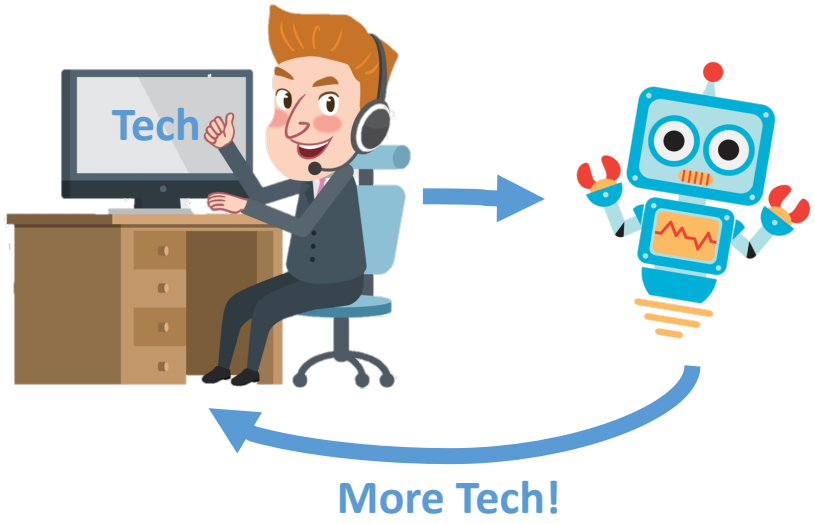
When Sebastian Thrun PERSON started at Google ORG in 2007 DATE , few people outside of the company took him seriously. “I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn’t worth talking to,” said Thrun PERSON , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE .



Named Entity Recognition (NER)

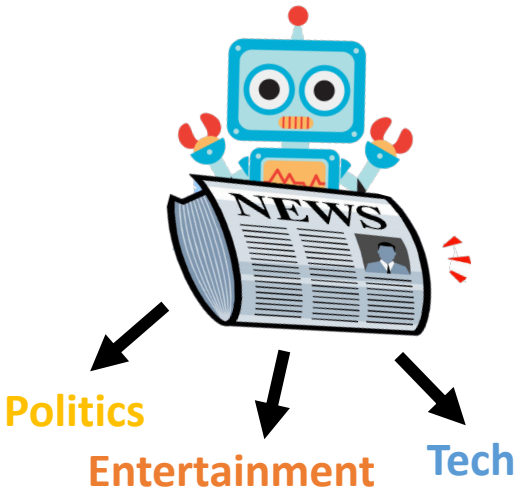
When Sebastian Thrun PERSON started at Google ORG in 2007 DATE , few people outside of the company took him seriously. “I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn’t worth talking to,” said Thrun PERSON , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE .

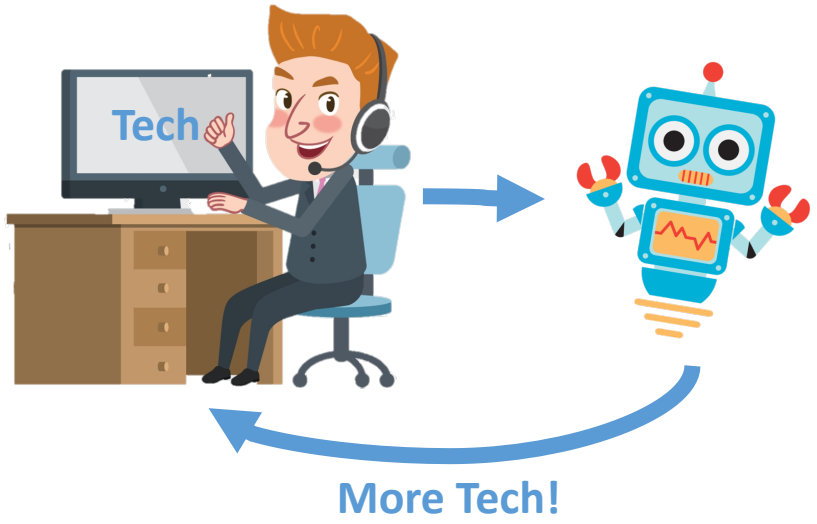


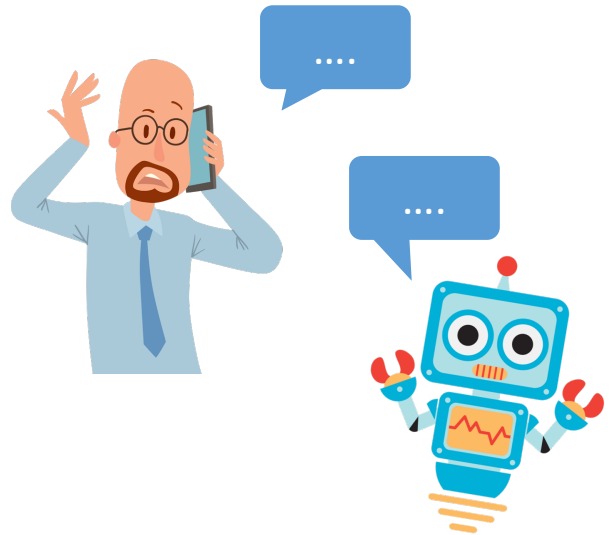


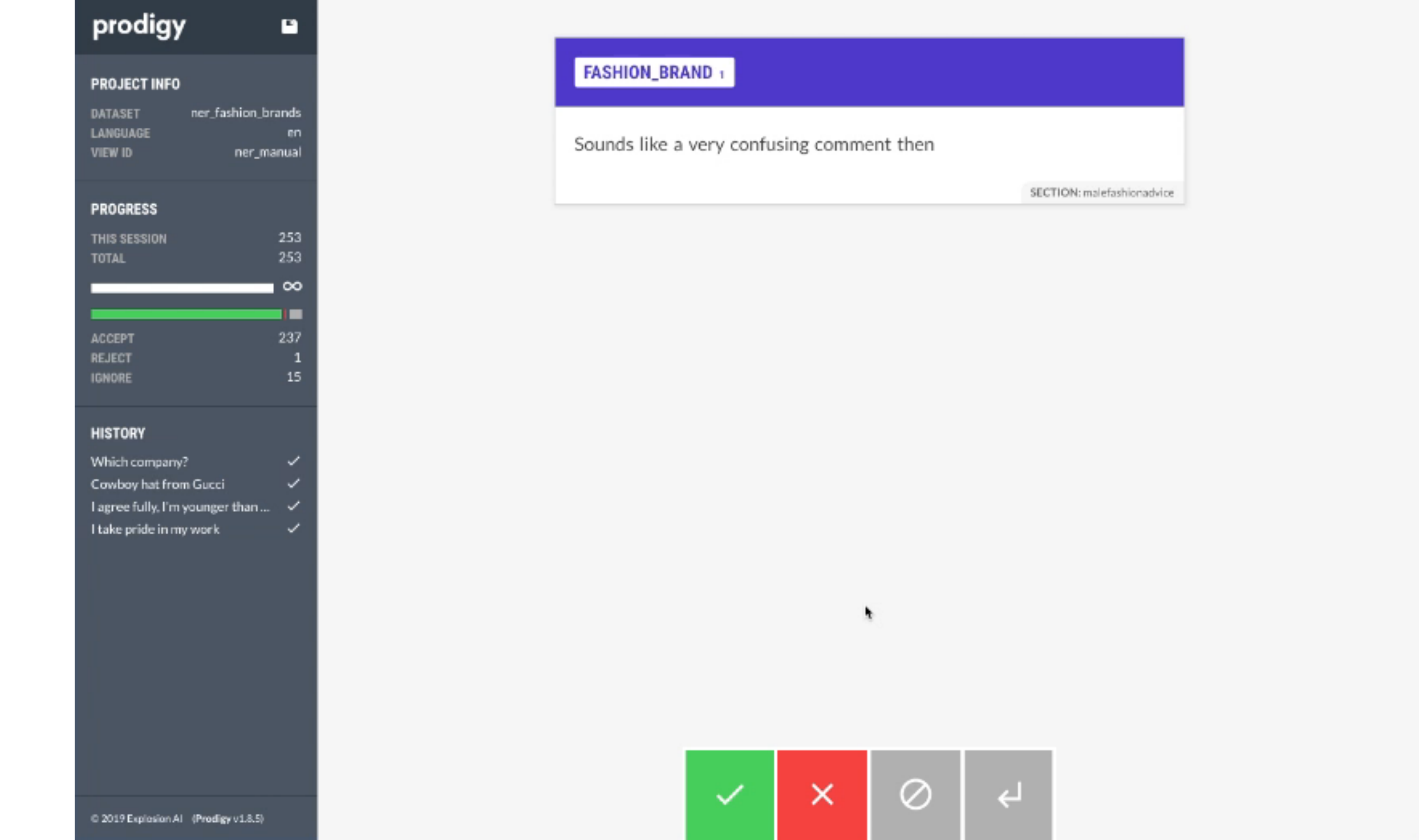
Named Entity Recognition (NER)

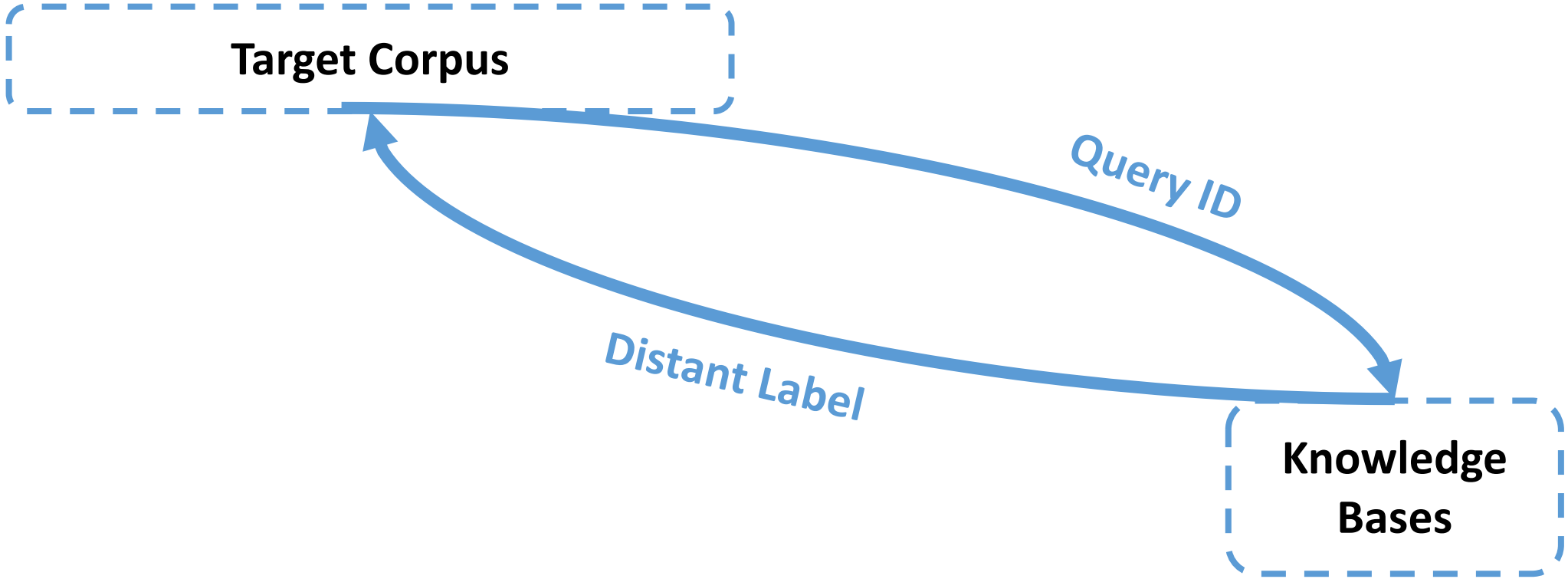
When Sebastian Thrun PERSON started at Google ORG in 2007 DATE , few people outside of the company took him seriously. “I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn’t worth talking to,” said Thrun PERSON , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE .



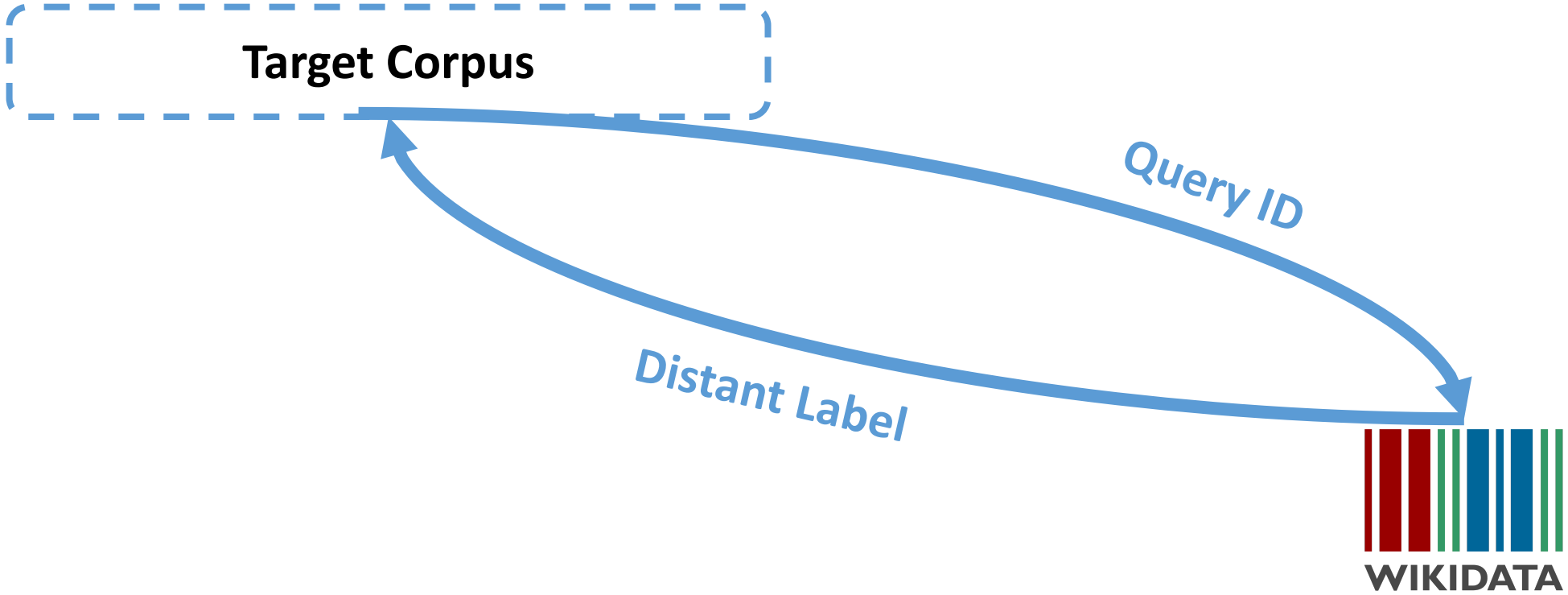




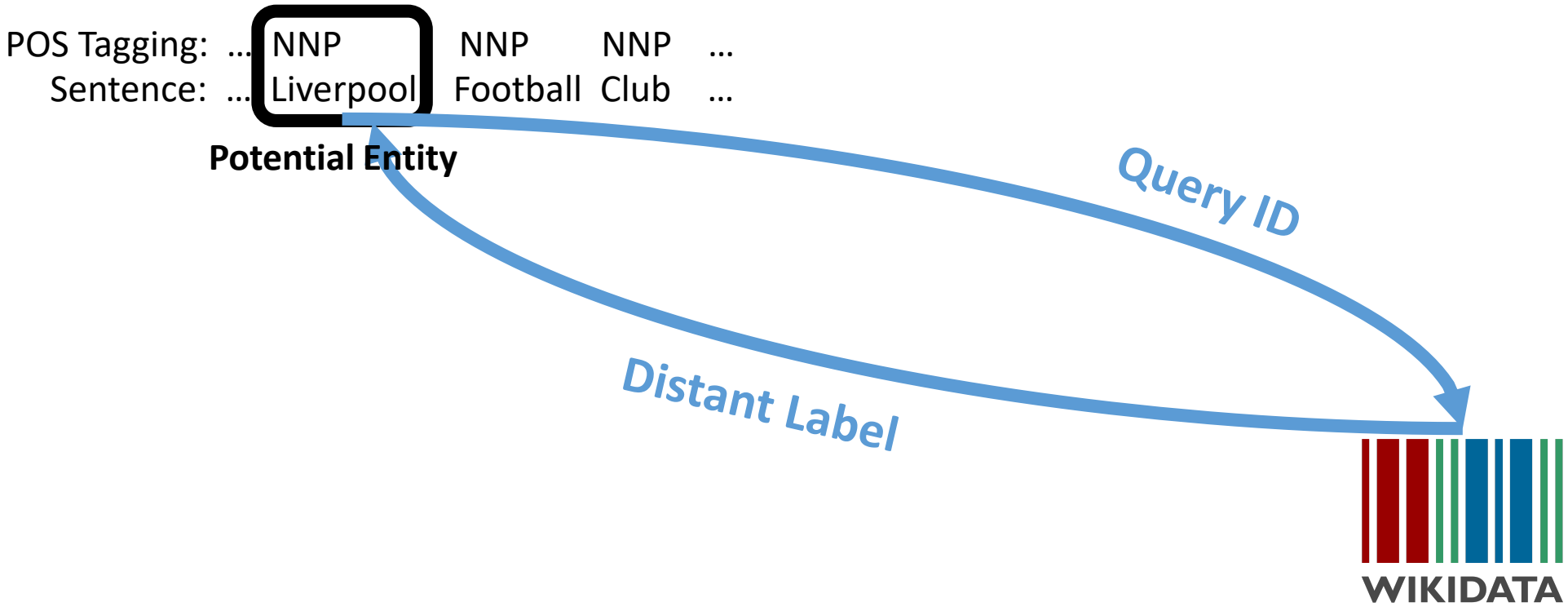

prodigy
PROJECT INFO
DATASET ner_fashion_brands
LANGUAGE en
VIEW ID ner_manual
PROGRESS
THIS SESSION 253
TOTAL 253
ACCEPT 237
REJECT 1
IGNORE 15
HISTORY
Which company?
Cowboy hat from Gucci
I agree fully. I'm younger than ...
I take pride in my work
© 2019 Explosion AI (Prodigy v1.8.5)
FASHION_BRAND 1
Sounds like a very confusing comment then
SECTION: malefashionadvice

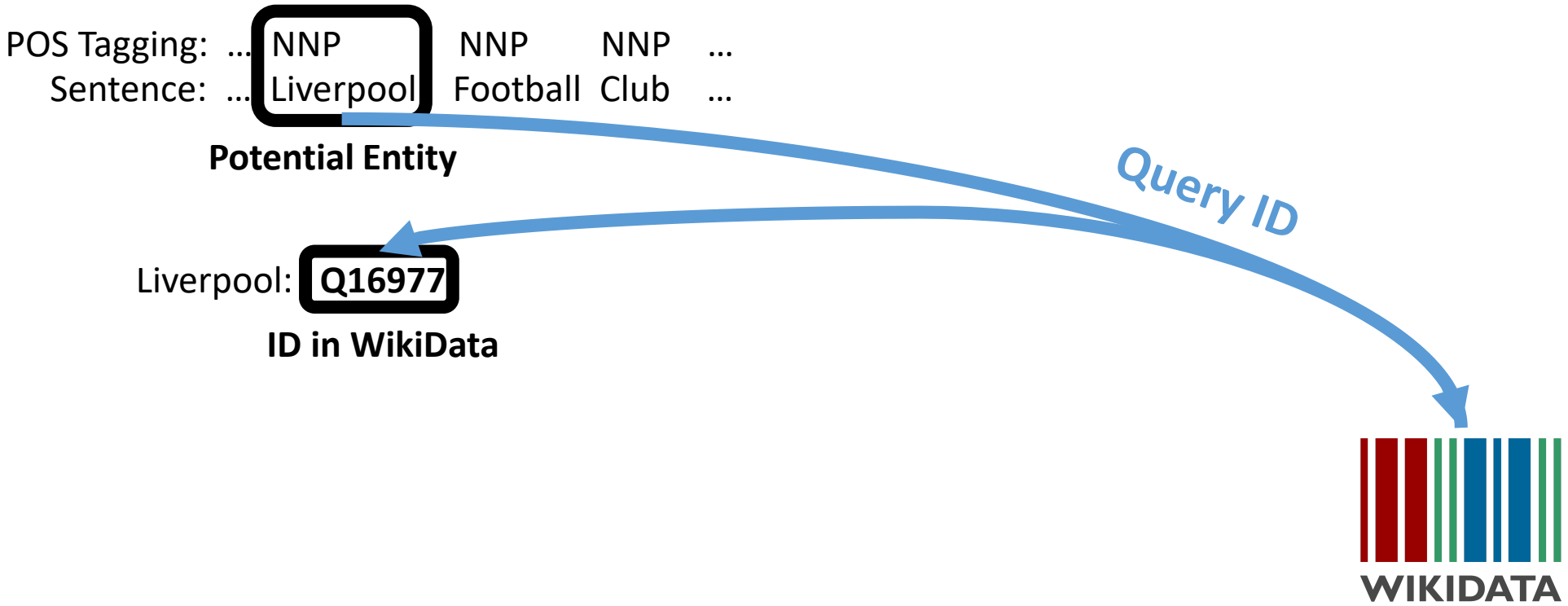


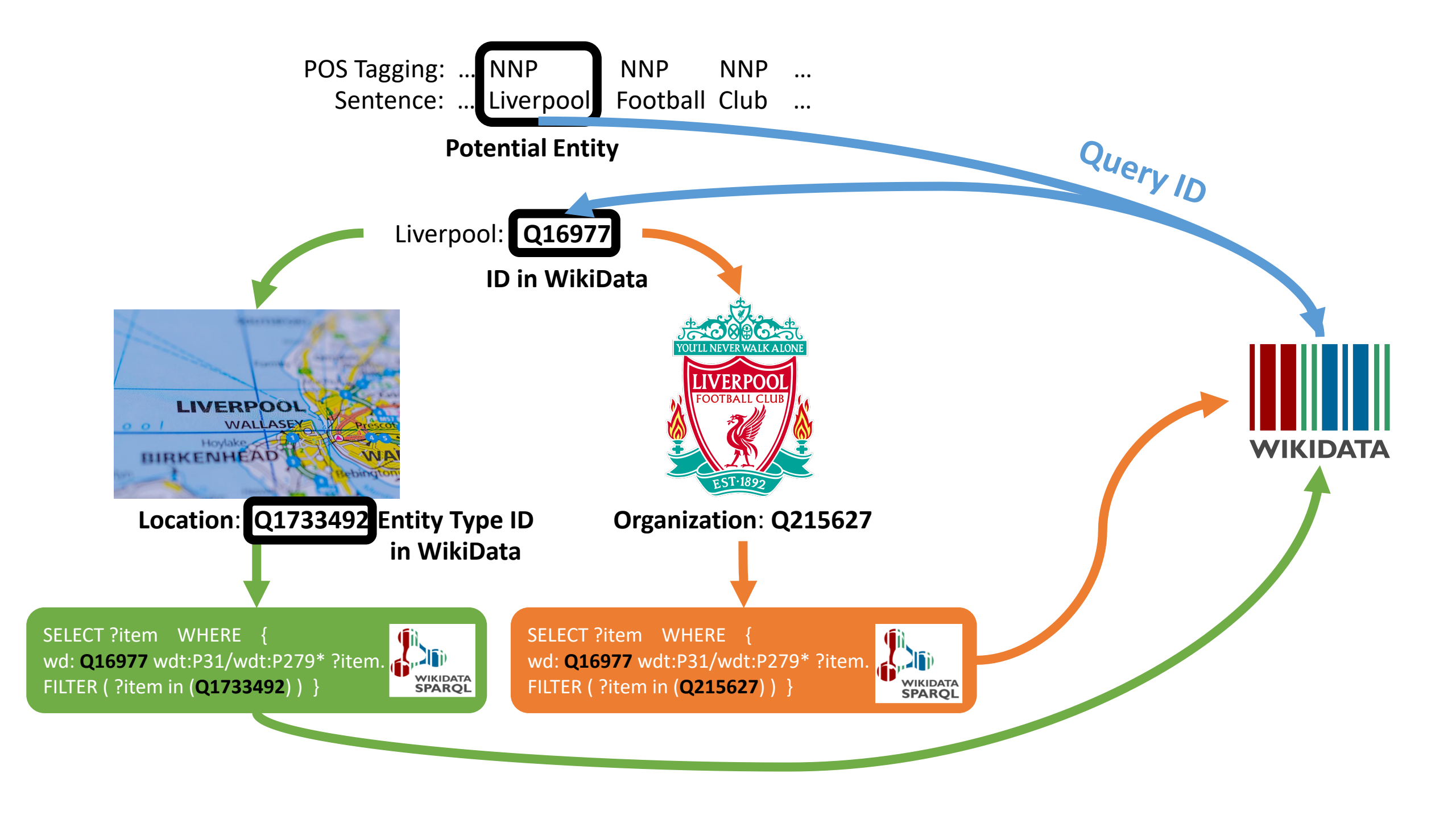
Distant Supervision

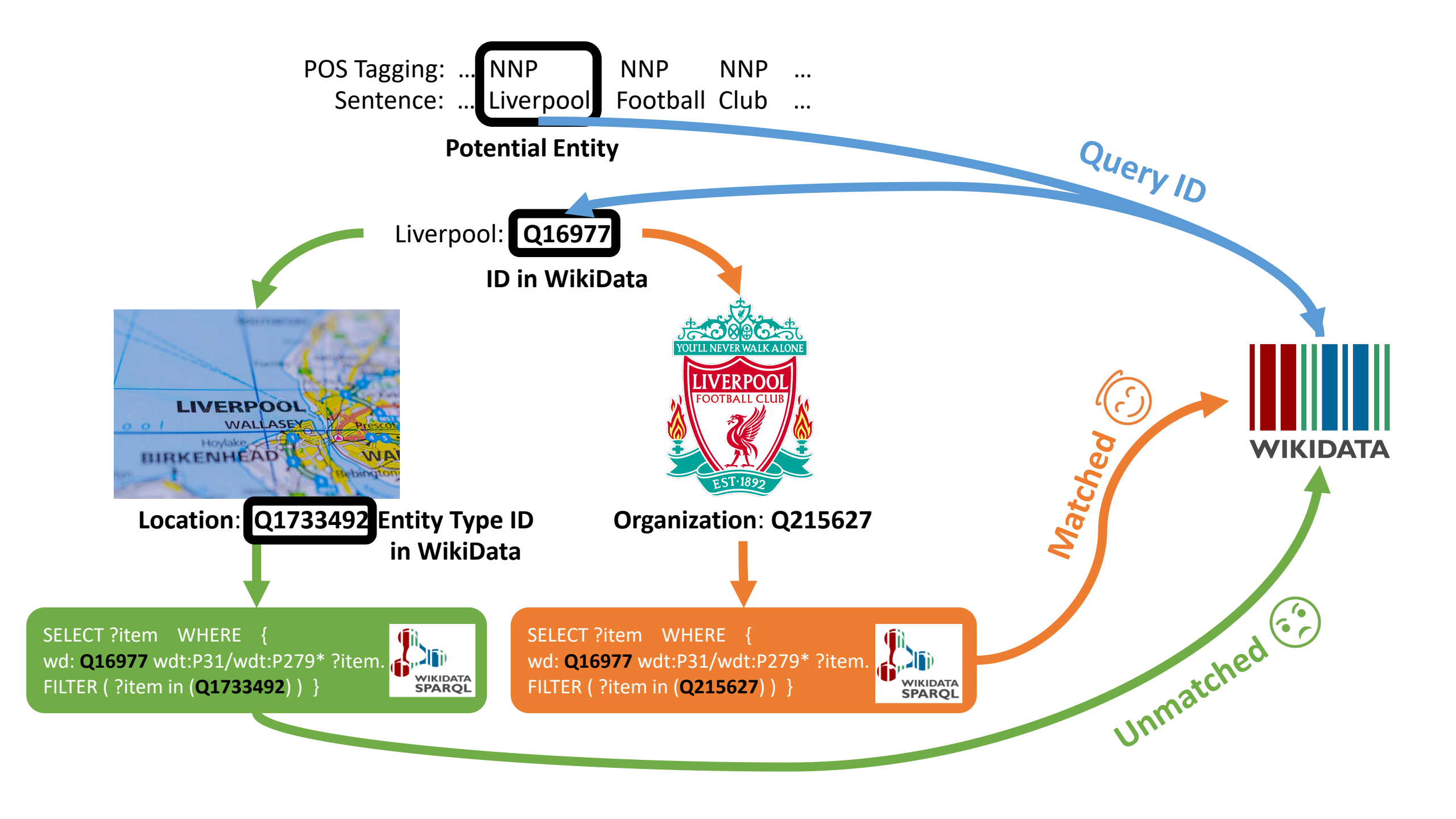


Distant Supervision


POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …
Potential Entity
Query ID
Distant Label
WIKIDATA


POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …
Potential Entity
Query ID
Liverpool: Q16977
ID in WikiData
WIKIDATA


POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …
Potential Entity
Query ID
Liverpool: Q16977
ID in WikiData
LIVERPOOL
WALLASEY
BIRKENHEAD
Location: Q1733492 Entity Type ID in WikiData
YOU'LL NEVER WALK ALONE
LIVERPOOL
FOOTBALL CLUB
EST·1892
Organization: Q215627
WIKIDATA
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER (?item in (Q1733492)) }
WIKIDATA SPARQL
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER (?item in (Q215627)) }
WIKIDATA SPARQL


POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …
Potential Entity
Query ID
Liverpool: Q16977
ID in WikiData
LIVERPOOL
WALLASEY
BIRKENHEAD
Location: Q1733492 Entity Type ID in WikiData
YOU'LL NEVER WALK ALONE
LIVERPOOL
FOOTBALL CLUB
EST·1892
Organization: Q215627
WIKIDATA
Matched
Unmatched
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER (?item in (Q1733492)) }
WIKIDATA SPARQL
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER (?item in (Q215627)) }
WIKIDATA SPARQL

POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …

Potential Entity

Noisy Annotation!

Query ID

Liverpool: **Q16977**

ID in WikiData

Location: Q1733492 Entity Type ID in WikiData

Organization: Q215627

```
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER ( ?item in (Q1733492) ) }
```

WIKIDATA SPARQL

```
SELECT ?item WHERE {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER ( ?item in (Q215627) ) }
```

WIKIDATA SPARQL

Matched

WIKIDATA

Unmatched

Matched!

POS Tagging: … NNP NNP NNP …
Sentence: … Liverpool Football Club …

Potential Entity

Incomplete Annotation!

Query ID

Liverpool: **Q16977**

ID in WikiData

Location: Q1733492 Entity Type ID in WikiData

Organization: Q215627

WIKIDATA

Matched

Unmatched

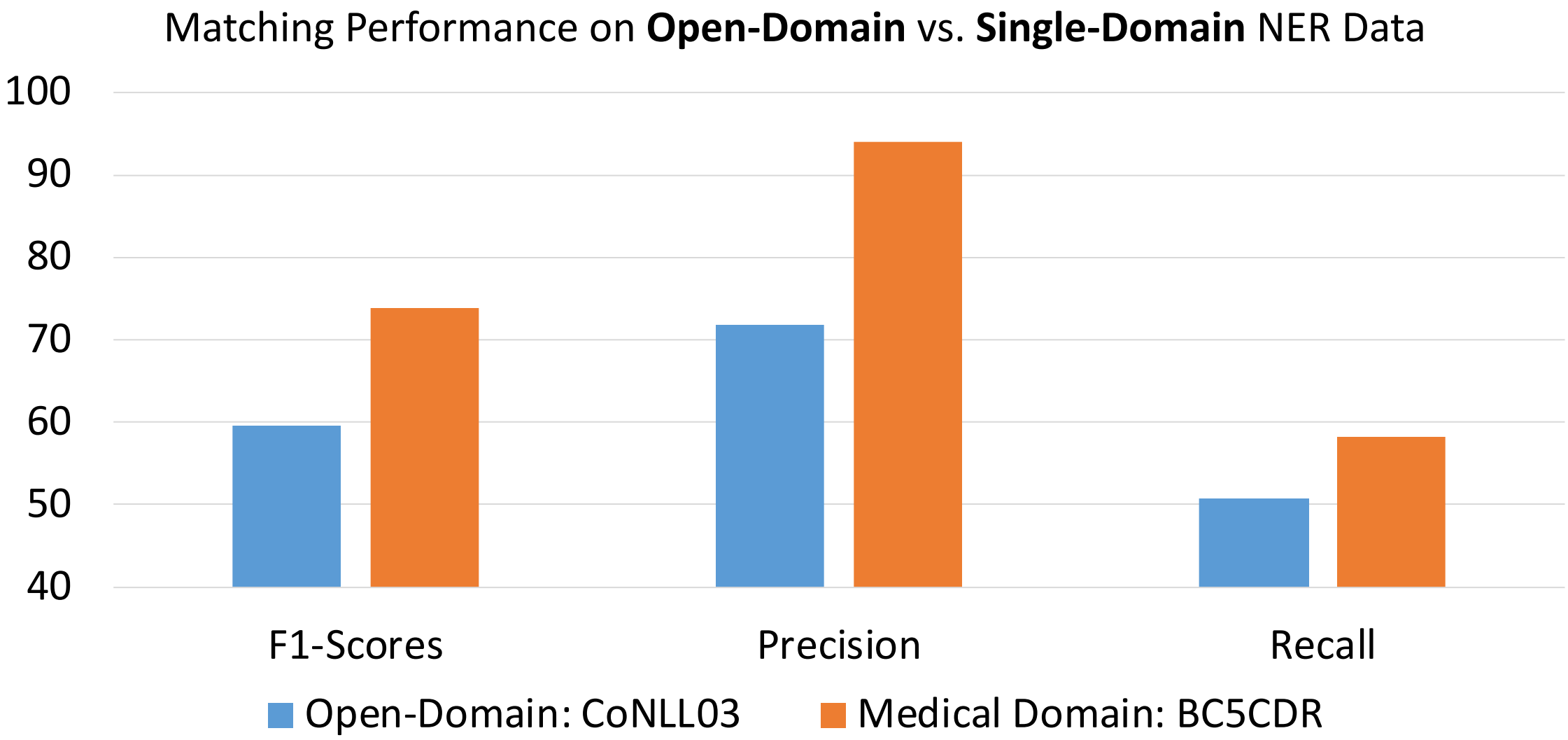
```
SELECT ?item  WHERE  {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER ( ?item in (Q1733492) ) }
```

WIKIDATA SPARQL

```
SELECT ?item  WHERE  {
wd: Q16977 wdt:P31/wdt:P279* ?item.
FILTER ( ?item in (Q215627) ) }
```

WIKIDATA SPARQL

Open Domain NER with Distant Supervision



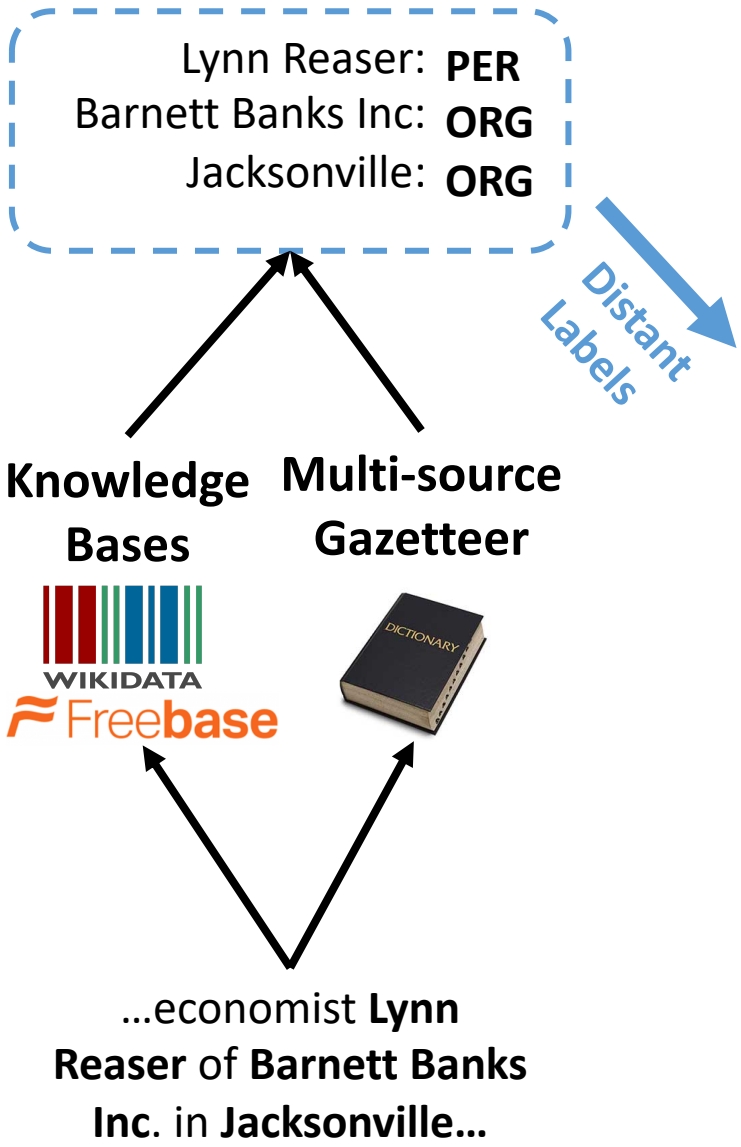
BOND: BERT-Assisted Open Domain NER with Distant Supervision

BOND: **B**ERT-Assisted **O**pen Domain **N**ER with **D**istant Supervision

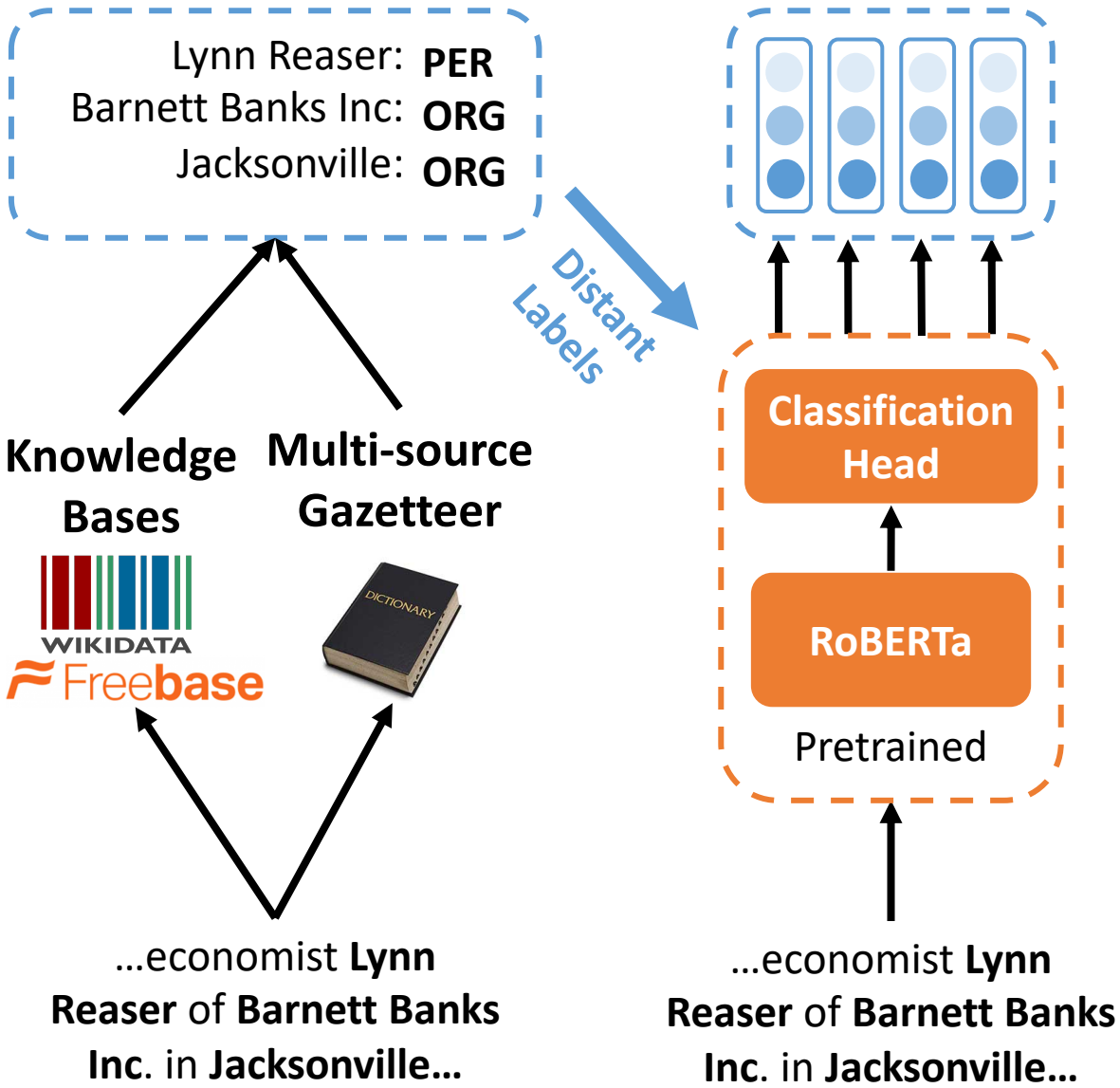
- Leverage power of **pre-trained language models** (e.g. BERT, RoBERTa)

BOND: **B**ERT-Assisted **O**pen Domain **N**ER with **D**istant Supervision

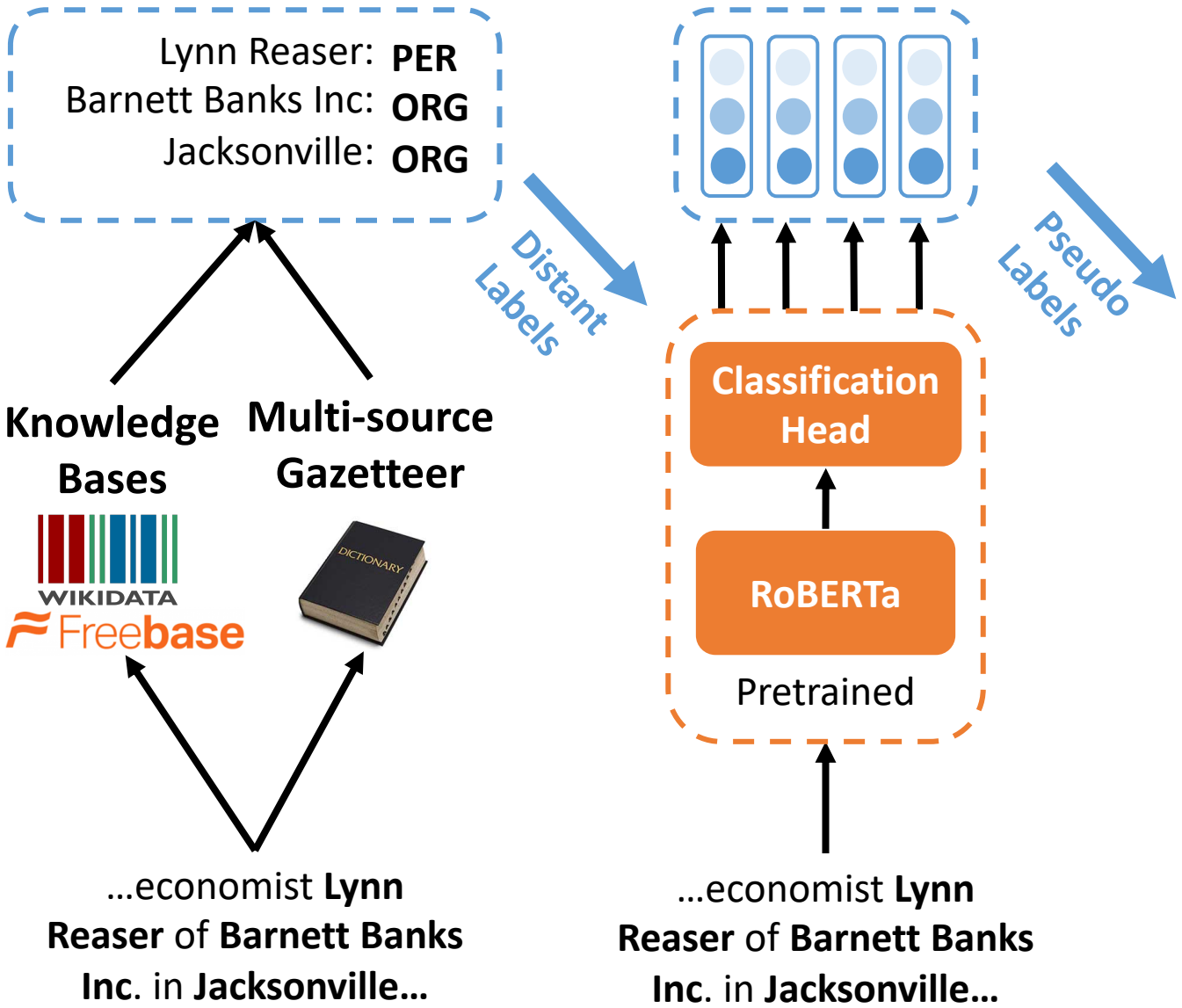
- Leverage power of **pre-trained language models** (e.g. RoBERTa)
- **Two-stage self-training framework**

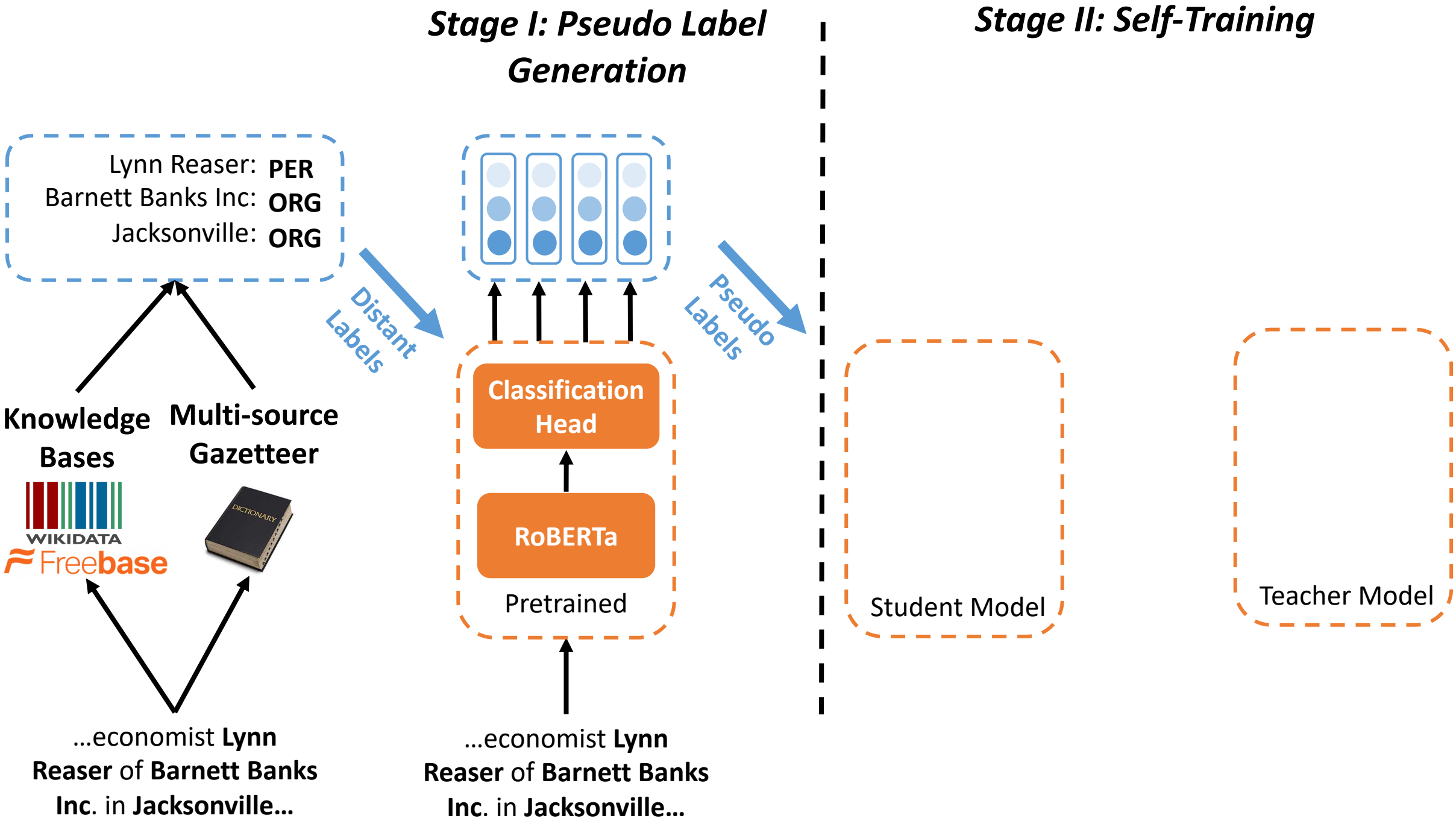

Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Distant Labels
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
…economist Lynn Reaser of Barnett Banks Inc. in Jacksonville…

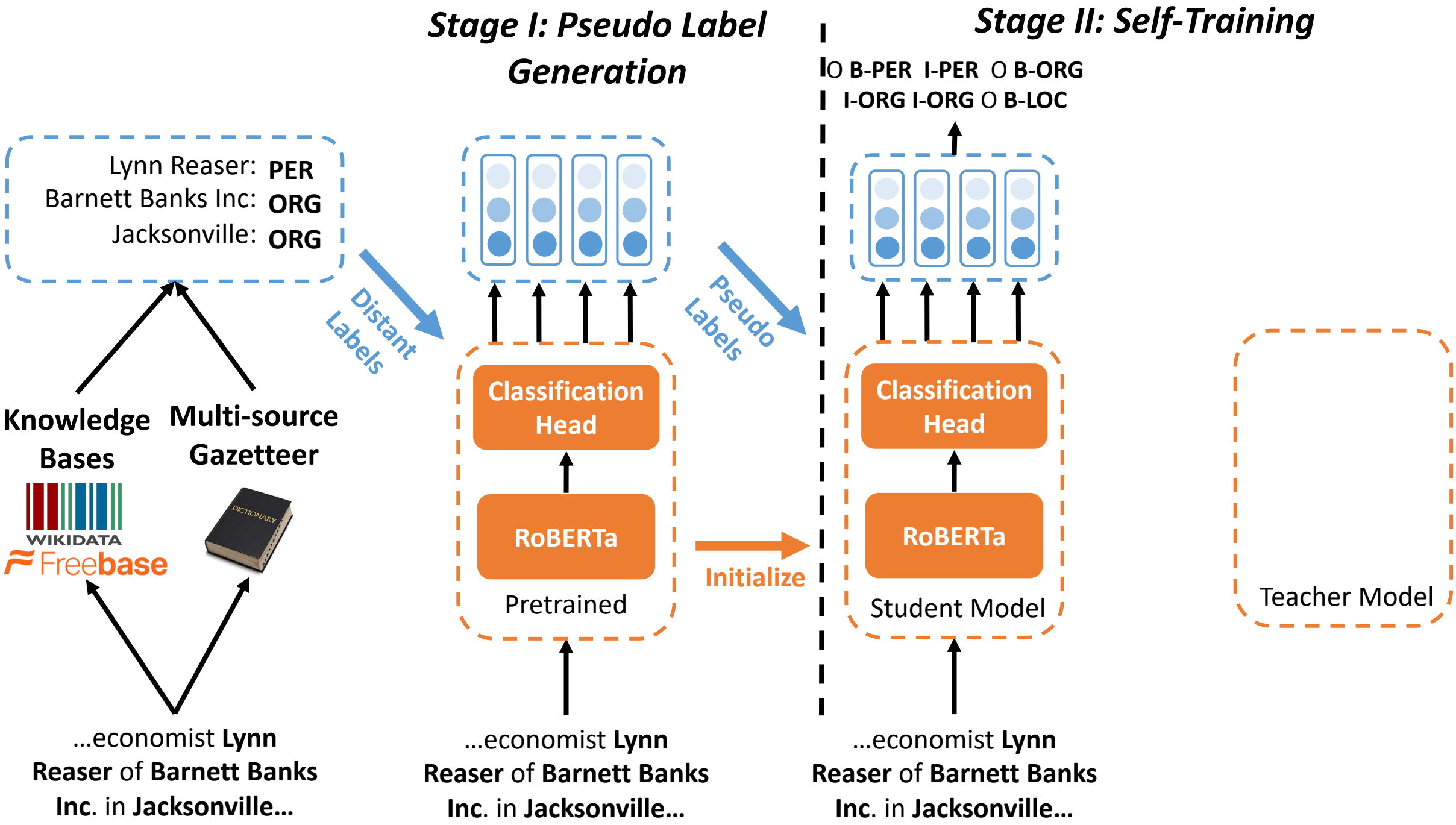
Stage I: Pseudo Label Generation

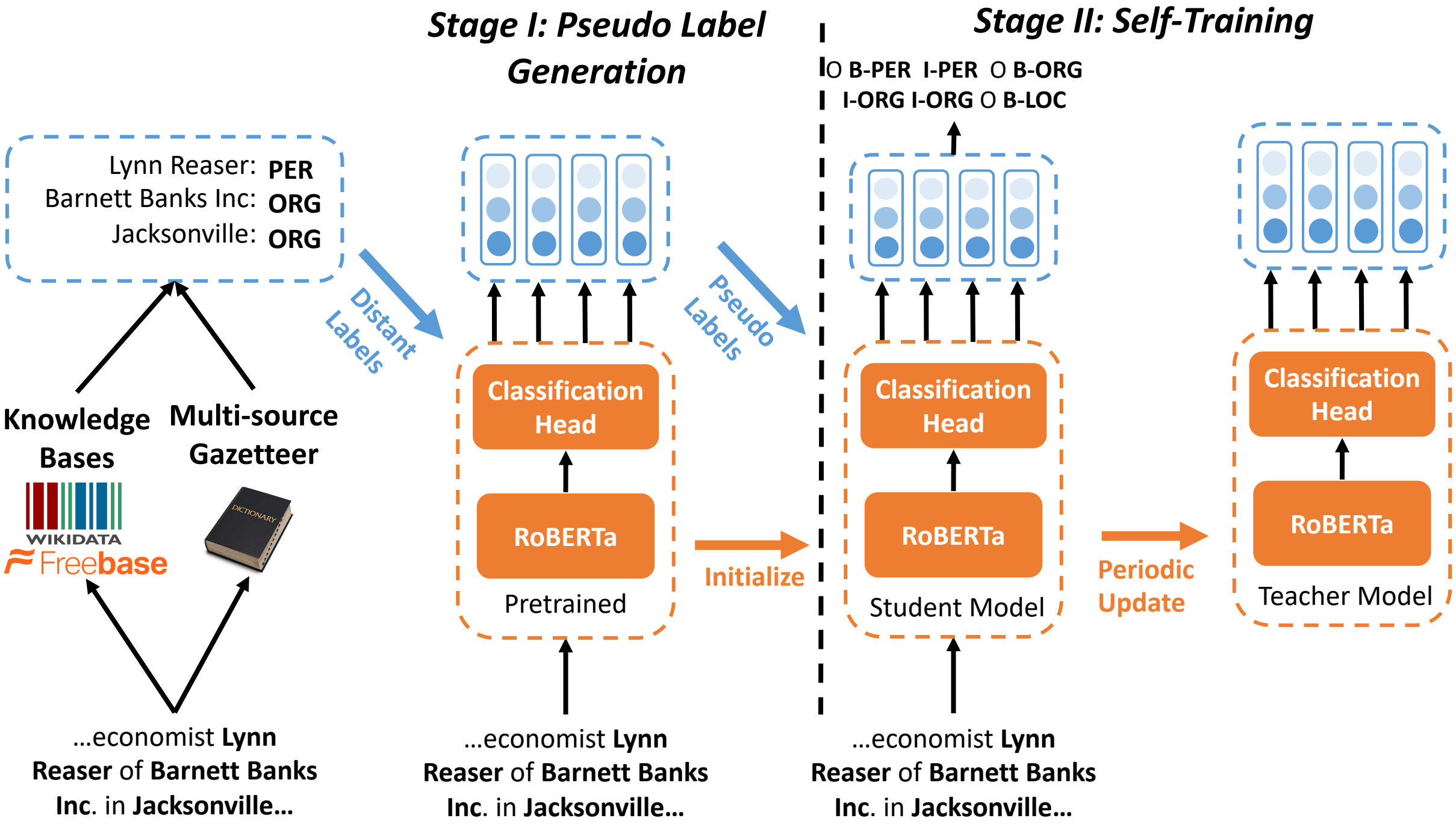


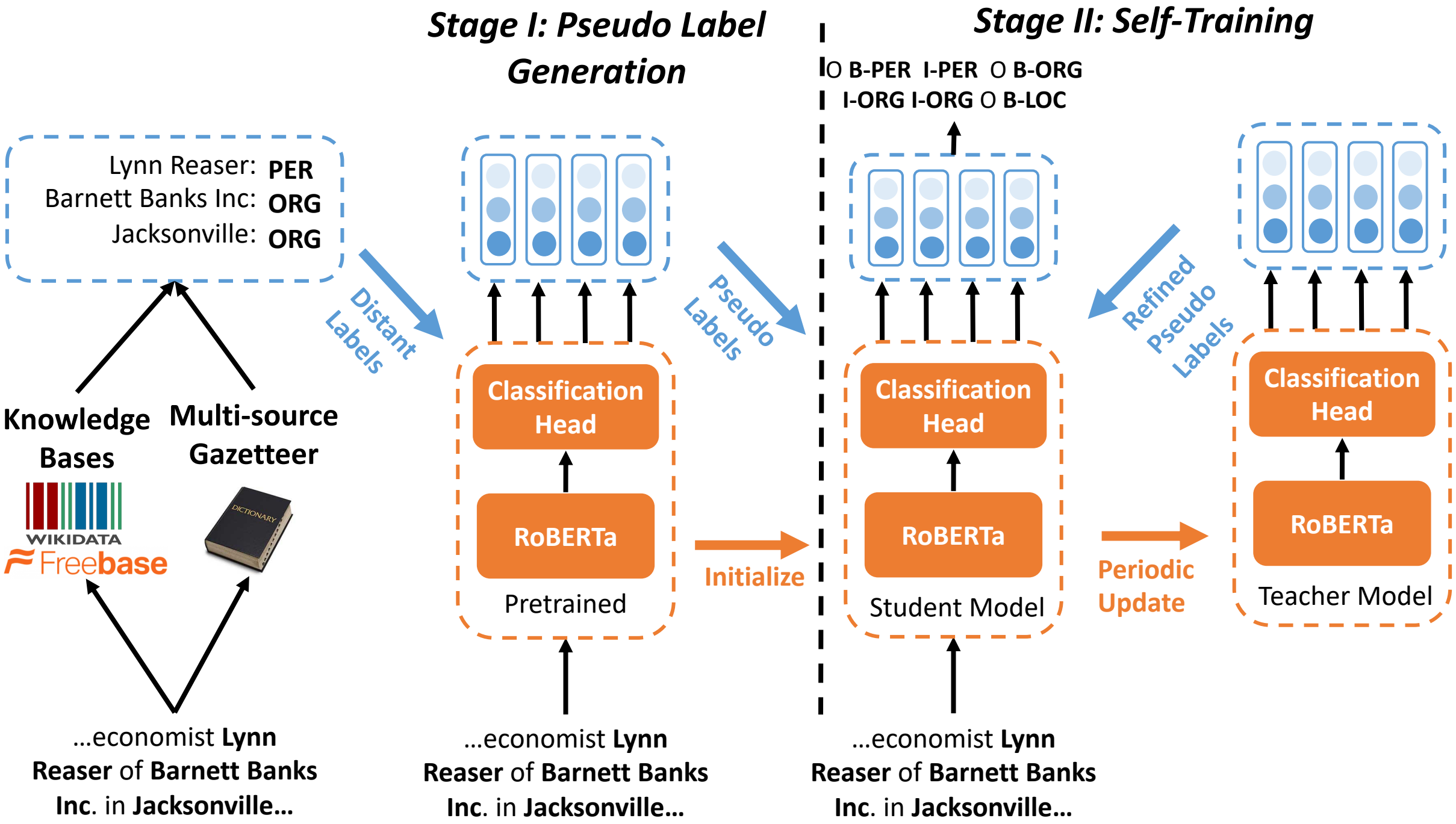
Stage I: Pseudo Label Generation

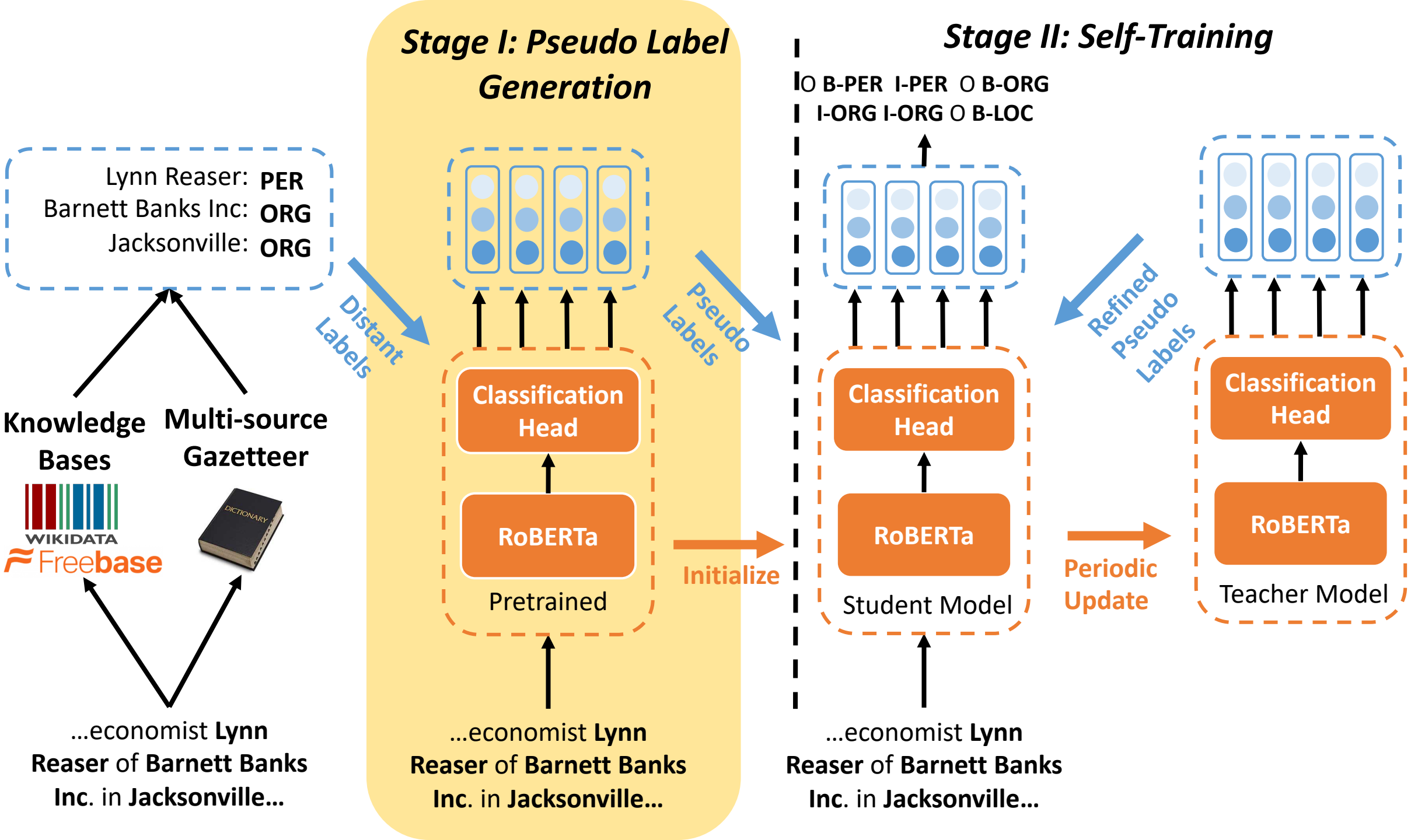



Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Distant Labels
Pseudo Labels
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
Classification Head
RoBERTa
Pretrained
Student Model
Teacher Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...

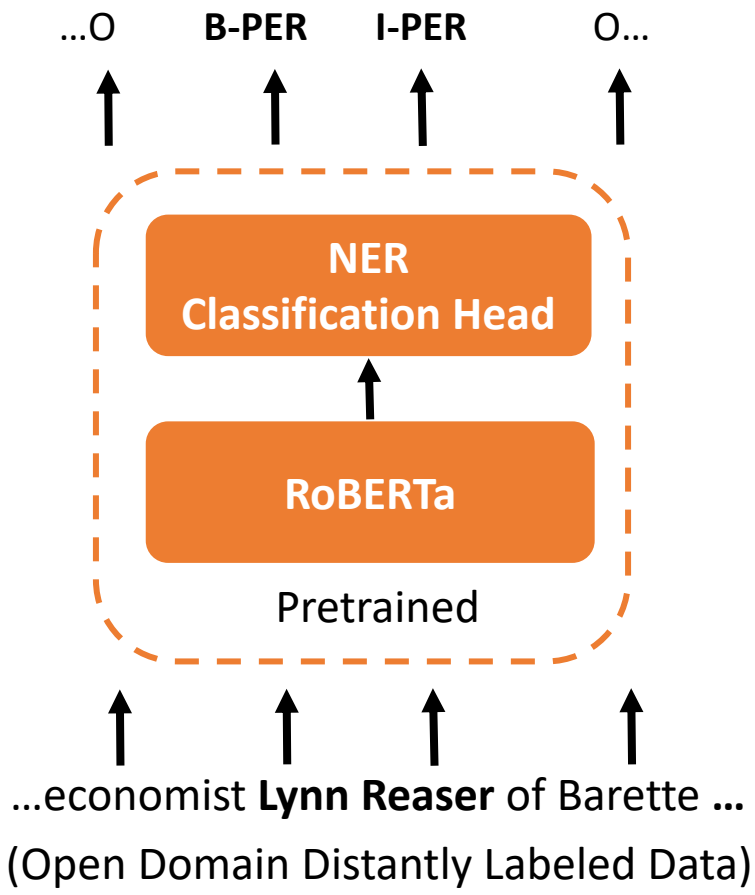

Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Distant Labels
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Classification Head
RoBERTa
Pretrained
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Pseudo Labels
O B-PER I-PER O B-ORG
I-ORG I-ORG O B-LOC
Initialize
Classification Head
RoBERTa
Student Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Teacher Model


Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
Distant Labels
Pseudo Labels
Classification Head
RoBERTa
Pretrained
Initialize
O B-PER I-PER O B-ORG
I-ORG I-ORG O B-LOC
Classification Head
RoBERTa
Student Model
Periodic Update
Classification Head
RoBERTa
Teacher Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...

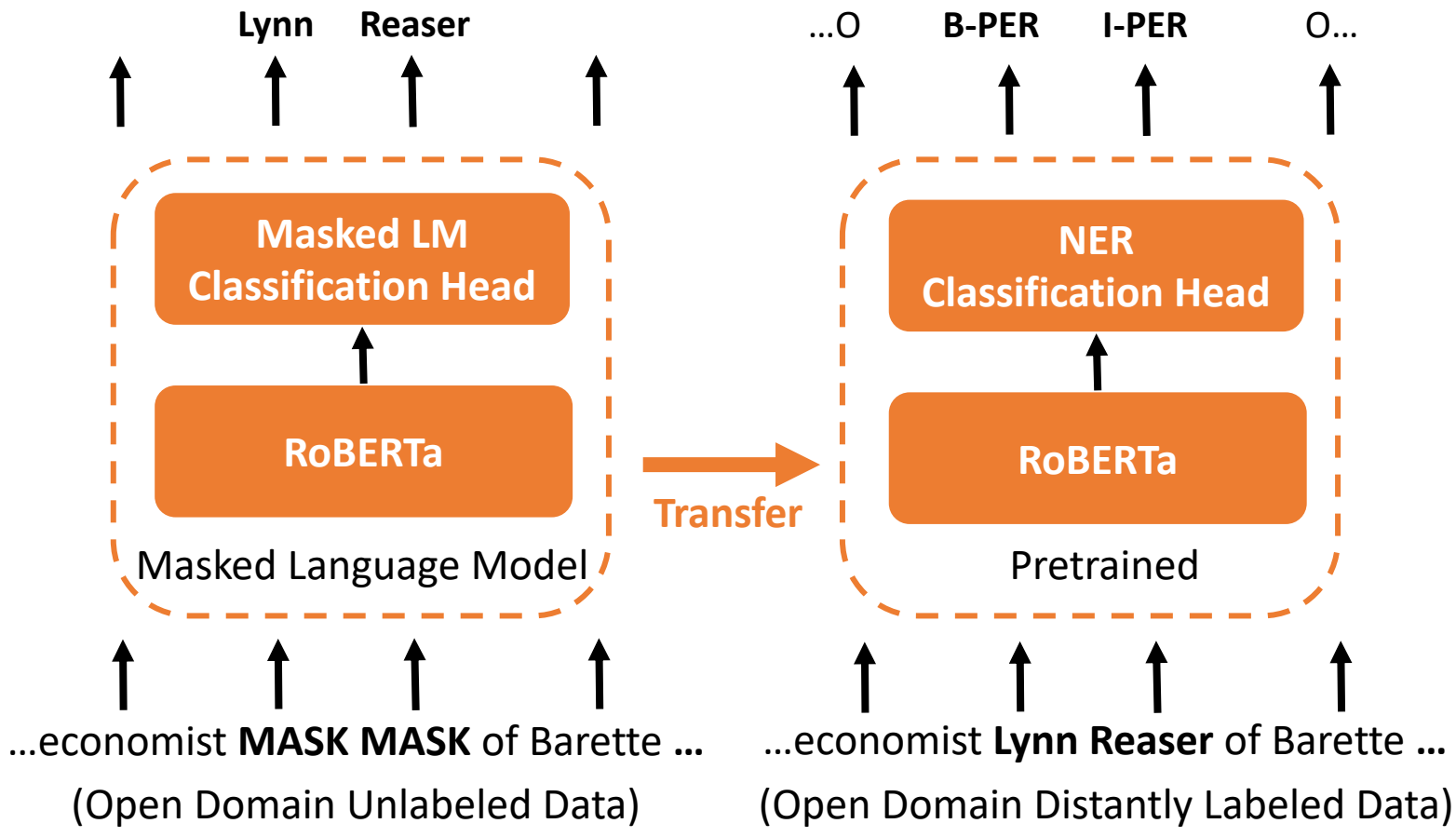

Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Distant Labels
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Classification Head
RoBERTa
Pretrained
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Pseudo Labels
Initialize
O B-PER I-PER O B-ORG
I-ORG I-ORG O B-LOC
Classification Head
RoBERTa
Student Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Refined Pseudo Labels
Periodic Update
Classification Head
RoBERTa
Teacher Model


Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Distant Labels
Classification Head
RoBERTa
Pretrained
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Pseudo Labels
Initialize
O B-PER I-PER O B-ORG
I-ORG I-ORG O B-LOC
Classification Head
RoBERTa
Student Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Refined Pseudo Labels
Periodic Update
Classification Head
RoBERTa
Teacher Model

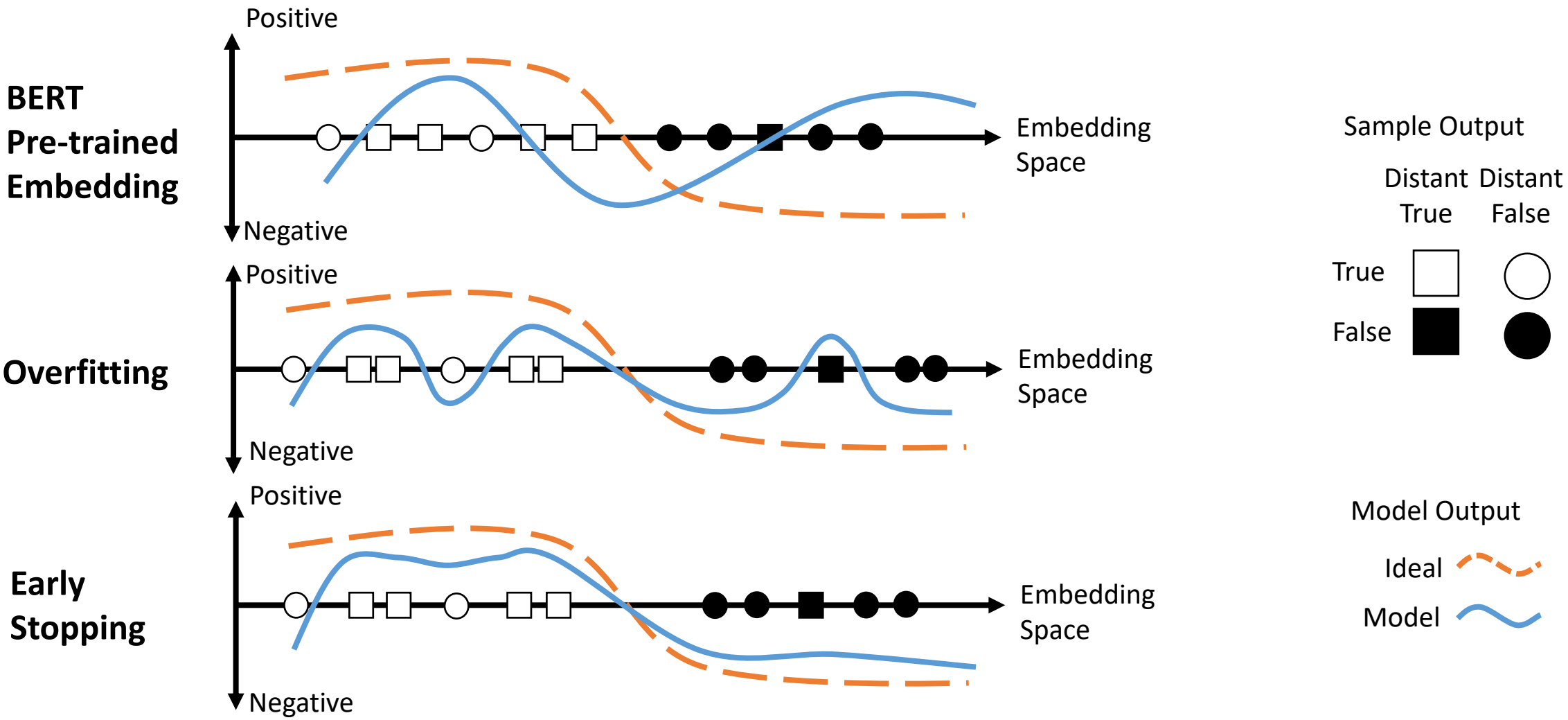
Stage I: BERT-Assisted Distantly Supervised Learning

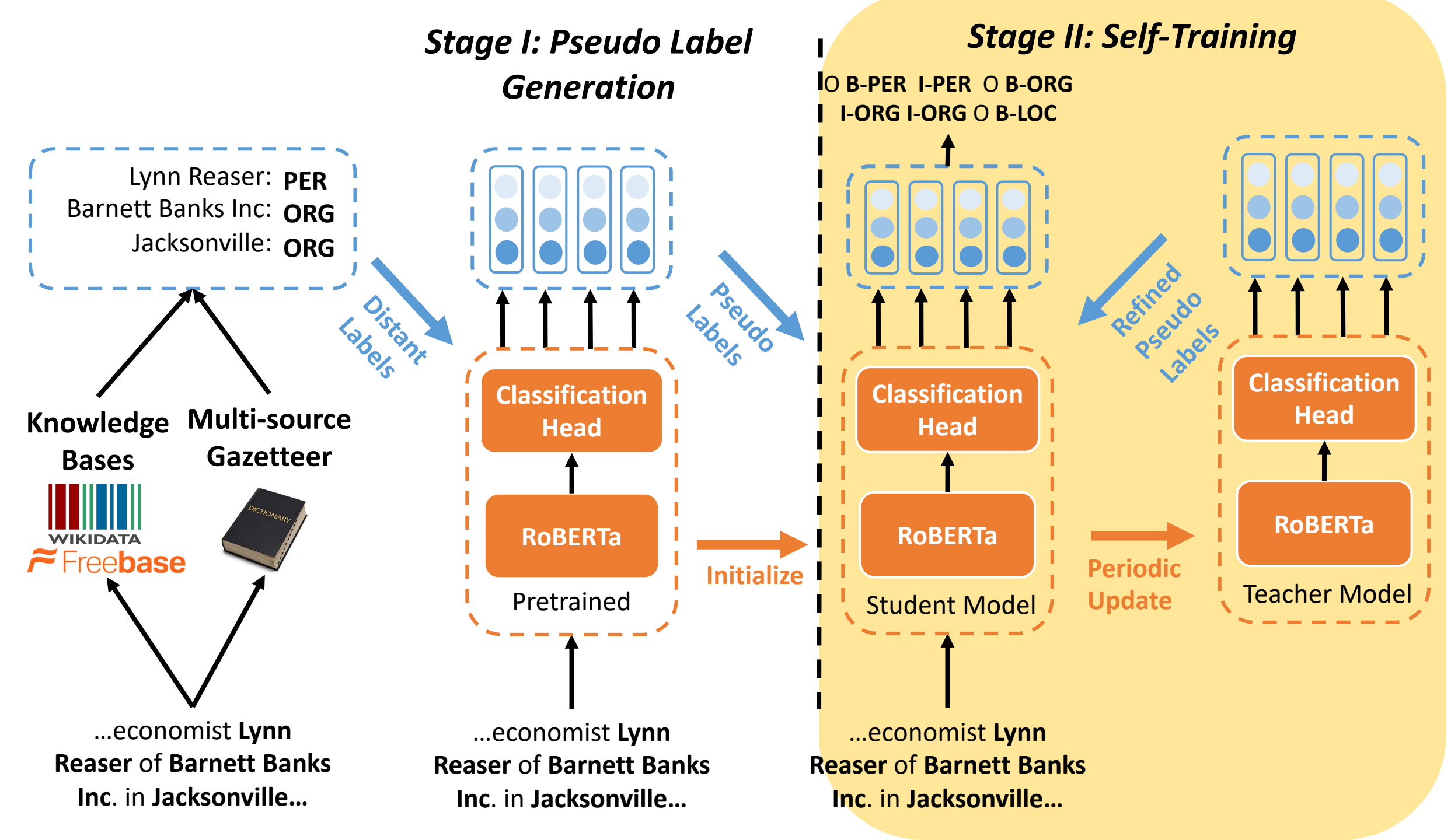


Stage I: BERT-Assisted Distantly Supervised Learning



Stage I: Early Stopping




Stage I: Pseudo Label Generation
Stage II: Self-Training
Lynn Reaser: PER
Barnett Banks Inc: ORG
Jacksonville: ORG
Distant Labels
Knowledge Bases
Multi-source Gazetteer
WIKIDATA
Freebase
DICTIONARY
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Classification Head
RoBERTa
Pretrained
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Pseudo Labels
Initialize
O B-PER I-PER O B-ORG
I-ORG I-ORG O B-LOC
Classification Head
RoBERTa
Student Model
...economist Lynn Reaser of Barnett Banks Inc. in Jacksonville...
Refined Pseudo Labels
Periodic Update
Classification Head
RoBERTa
Teacher Model

Stage II: Teacher-Student Framework

- Initialize the teacher model $f(\cdot;\theta_{tea})$ and the student model $f(\cdot;\theta_{stu})$ with the early stopped model $f(\cdot;\hat{\theta})$ obtained in Stage I

$$\theta_{tea}^{(0)} = \theta_{stu}^{(0)} = \hat{\theta}$$

Stage II: Teacher-Student Framework

- Initialize the teacher model $f(\cdot;\theta_{tea})$ and the student model $f(\cdot;\theta_{stu})$ with the early stopped model $f(\cdot;\hat{\theta})$ obtained in Stage I

$$\theta_{tea}^{(0)} = \theta_{stu}^{(0)} = \hat{\theta}$$

- At the t-th iteration
  - The teacher model generates pseudo labeled data $\{X_m, \tilde{Y}_m\}_{m=1}^{M}$

$$\tilde{y}_{m,n}^{(t)} = \underset{c}{\operatorname{argmax}}\ f_{n,c}(X_m;\theta_{tea}^{(t)})$$

Stage II: Teacher-Student Framework

- Initialize the teacher model $f(\cdot;\theta_{tea})$ and the student model $f(\cdot;\theta_{stu})$ with the early stopped model $f(\cdot;\hat{\theta})$ obtained in Stage I

$$\theta_{tea}^{(0)} = \theta_{stu}^{(0)} = \hat{\theta}$$

- At the t-th iteration
  - The teacher model generates pseudo labeled data $\{X_m, \tilde{Y}_m\}_{m=1}^{M}$

$$\tilde{y}_{m,n}^{(t)} = \underset{c}{\operatorname{argmax}}\ f_{n,c}(X_m;\theta_{tea}^{(t)})$$

  - The student model fits these pseudo-labels by solving

$$\theta_{stu}^{(t)} = \underset{\theta}{\operatorname{argmin}} \frac{1}{M}\sum_{m=1}^{M} \ell(\tilde{Y}_m^{(t)}, f(X_m;\theta))$$

Stage II: Teacher-Student Framework

- Initialize the teacher model $f(\cdot;\theta_{tea})$ and the student model $f(\cdot;\theta_{stu})$ with the early stopped model $f(\cdot;\hat{\theta})$ obtained in Stage I

$$\theta_{tea}^{(0)} = \theta_{stu}^{(0)} = \hat{\theta}$$

- At the t-th iteration
  - The teacher model generates pseudo labeled data $\{X_m, \tilde{Y}_m\}_{m=1}^{M}$

$$\tilde{y}_{m,n}^{(t)} = \underset{c}{\operatorname{argmax}}\ f_{n,c}(X_m;\theta_{tea}^{(t)})$$

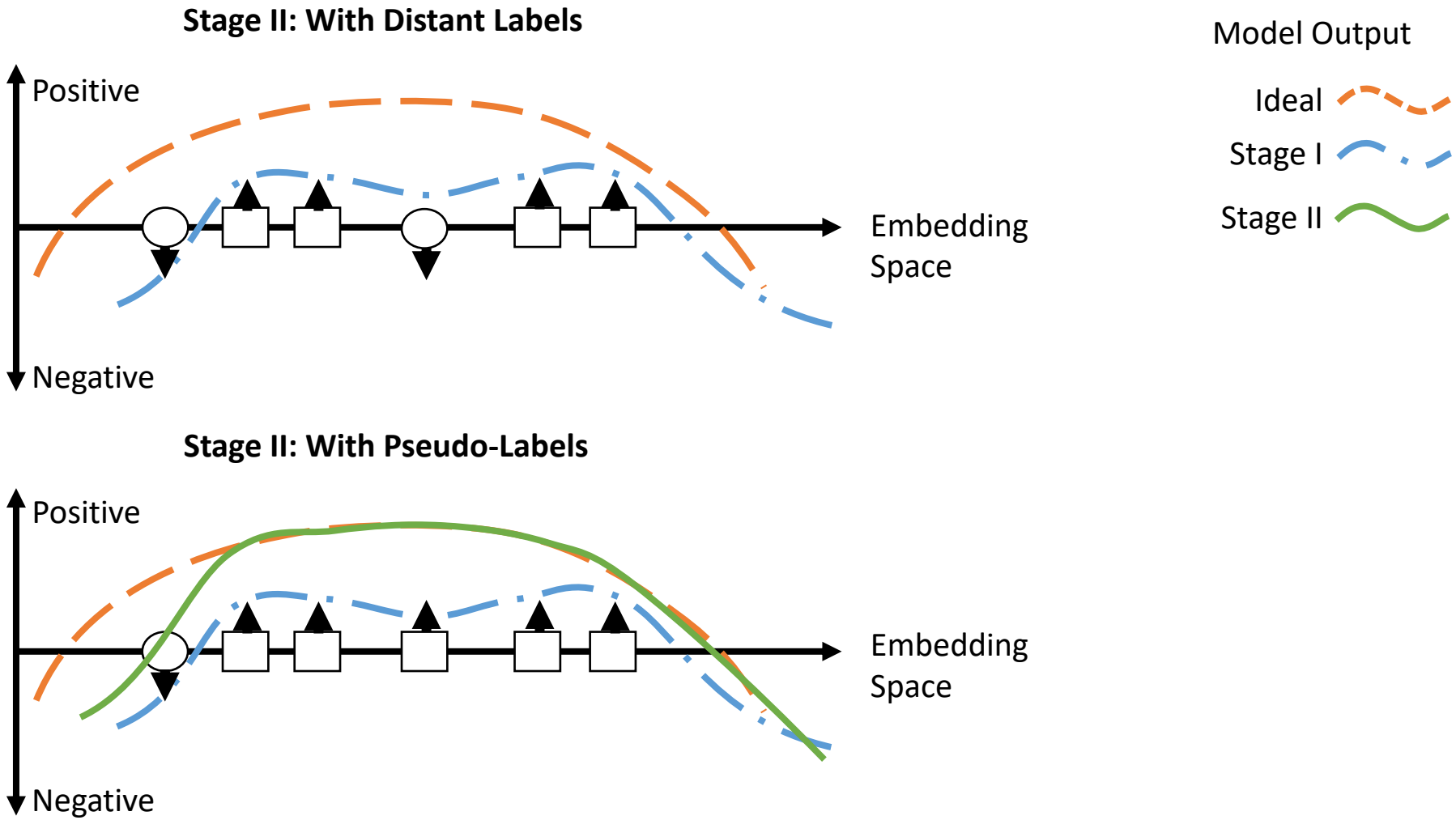
  - The student model fits these pseudo-labels by solving

$$\theta_{stu}^{(t)} = \underset{\theta}{\operatorname{argmin}} \frac{1}{M}\sum_{m=1}^{M} \ell(\tilde{Y}_m^{(t)}, f(X_m;\theta))$$

  - Update the teacher model and the student model by

$$\theta_{tea}^{(t+1)} = \theta_{stu}^{(t+1)} = \hat{\theta}_{stu}^{(t)}$$

Stage II: Teacher-Student Framework



Stage II: Soft Labels w/ Confidence Re-weighting

- At the t-th iteration
  - Denote the output probability simplex over C classes as $[f_{n,c}(X_m; \theta)]_{c=1}^{C}$

Stage II: Soft Labels w/ Confidence Re-weighting

- At the t-th iteration
  - Denote the output probability simplex over C classes as $[f_{n,c}(X_m;\theta)]_{c=1}^{C}$
  - The teacher model compute the unnormalized frequency of the tokens belonging to the c-th class as

$$p_c = \sum_{m=1}^{M} \sum_{n=1}^{N} f_{n,c}(X_m;\theta_{tea}^{(t)})$$

Stage II: Soft Labels w/ Confidence Re-weighting

- At the t-th iteration
  - Denote the output probability simplex over C classes as $[f_{n,c}(X_m;\theta)]_{c=1}^{C}$
  - The teacher model compute the unnormalized frequency of the tokens belonging to the c-th class as

$$p_c = \sum_{m=1}^{M}\sum_{n=1}^{N} f_{n,c}(X_m;\theta_{tea}^{(t)})$$

  - The teacher model generates soft labels $\left\{S_m^{(t)} = \left[s_{m,n}^{(t)}\right]_{n=1}^{N}\right\}_{m=1}^{M}$ by

$$s_{m,n}^{(t)} = \left[s_{m,n,c}^{(t)}\right]_{c=1}^{C} = \left[\frac{f_{n,c}^2(X_m;\theta_{tea}^{(t)})/p_c}{\sum_{c'=1}^{C} f_{n,c'}^2(X_m;\theta_{tea}^{(t)})/p_{c'}}\right]_{c=1}^{C}$$

Stage II: Soft Labels w/ Confidence Re-weighting

- At the t-th iteration
  - Denote the output probability simplex over C classes as $[f_{n,c}(X_m;\theta)]_{c=1}^{C}$
  - The teacher model compute the unnormalized frequency of the tokens belonging to the c-th class as

$$p_c = \sum_{m=1}^{M}\sum_{n=1}^{N} f_{n,c}(X_m;\theta_{tea}^{(t)})$$

  - The teacher model generates soft labels $\left\{S_m^{(t)} = \left[s_{m,n}^{(t)}\right]_{n=1}^{N}\right\}_{m=1}^{M}$ by

$$s_{m,n}^{(t)} = \left[s_{m,n,c}^{(t)}\right]_{c=1}^{C} = \left[\frac{f_{n,c}^2(X_m;\theta_{tea}^{(t)})/p_c}{\sum_{c'=1}^{C} f_{n,c'}^2(X_m;\theta_{tea}^{(t)})/p_{c'}}\right]_{c=1}^{C}$$

  - The student model fit the soft labels by solving

$$\theta_{stu}^{(t)} = \underset{\theta}{\operatorname{argmin}} \frac{1}{M}\sum_{m=1}^{M} \ell_{KL}(S_m^{(t)}, f(X_m;\theta)),$$

where $\ell_{KL}(S_m^{(t)}, f(X_m;\theta)) = \frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} -s_{m,n,c}^{(t)} \log f_{n,c}(X_m;\theta)$

Stage II: High-Confidence Selection

- At the t-th iteration
  - Select a set of high confidence tokens from the m-th sentence by

$$H_m^{(t)} = \left\{ n : \max_c s_{m,n,c}^{(t)} > \epsilon \right\},$$

where $\epsilon \in (0,1)$

Stage II: High-Confidence Selection

- At the t-th iteration
  - Select a set of high confidence tokens from the m-th sentence by

$$H_m^{(t)} = \left\{ n : \max_c s_{m,n,c}^{(t)} > \epsilon \right\},$$

where $\epsilon \in (0,1)$

  - The student model fit the high-confidence labels of the selected tokens by solving

$$\theta_{stu}^{(t)} = \operatorname*{argmin}_{\theta} \frac{1}{M|H_m^{(t)}|} \sum_{m=1}^{M} \sum_{n \in H_m^{(t)}} -s_{m,n,c}^{(t)} \log f_{n,c}(X_m; \theta)$$

Experiment: Main Result

Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Tweet	OntoNote5.0	Webpage	Wikigold
Entity Types	4	10	18	4	4
KB Matching	71.40(81.13/63.75)	35.83(40.34/32.22)	59.51(63.86/55.71)	52.45(62.59/45.14)	47.76(47.90/47.63)
Fully-Supervised (Our implementation)					
RoBERTa	90.11(89.14/91.10)	52.19(51.76/52.63)	86.20(84.59/87.88)	72.39(66.29/79.73)	86.43(85.33/87.56)
BiLSTM-CRF	91.21(91.35/91.06)	52.18(60.01/46.16)	86.17(85.99/86.36)	52.34(50.07/54.76)	54.90(55.40/54.30)
Baseline (Our implementation)					
BiLSTM-CRF	59.50(75.50/49.10)	21.77(46.91/14.18)	66.41(68.44/64.50)	43.34(58.05/34.59)	42.92(47.55/39.11)
AutoNER	67.00(75.21/60.40)	26.10(43.26/18.69)	67.18(64.63/69.95)	51.39(48.82/54.23)	47.54(43.54/52.35)
LRNT	69.74(79.91/61.87)	23.84(46.94/15.98)	67.69(67.36/68.02)	47.74(46.70/48.83)	46.21(45.60/46.84)
Other Baseline (Reported Results)					
KALM †	76.00(– / –)	–	–	–	–
ConNET ◇	75.57(84.11/68.61)	–	–	–	–
Our BOND **Framework**					
Stage I	75.61(83.76/68.90)	46.61(53.11/41.52)	68.11(66.71/69.56)	59.11(60.14/58.11)	51.55(49.17/54.50)
BOND	81.48(82.05/80.92)	48.01(53.16/43.76)	68.35(67.14/69.61)	65.74(67.37/64.19)	60.07(53.44/68.58)

Experiment: Main Result

Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Tweet	OntoNote5.0	Webpage	Wikigold
Entity Types	4	10	18	4	4
KB Matching	71.40(81.13/63.75)	35.83(40.34/32.22)	59.51(63.86/55.71)	52.45(62.59/45.14)	47.76(47.90/47.63)
Fully-Supervised (Our implementation)					
RoBERTa	90.11(89.14/91.10)	52.19(51.76/52.63)	86.20(84.59/87.88)	72.39(66.29/79.73)	86.43(85.33/87.56)
BiLSTM-CRF	91.21(91.35/91.06)	52.18(60.01/46.16)	86.17(85.99/86.36)	52.34(50.07/54.76)	54.90(55.40/54.30)
Baseline (Our implementation)					
BiLSTM-CRF	59.50(75.50/49.10)	21.77(46.91/14.18)	66.41(68.44/64.50)	43.34(58.05/34.59)	42.92(47.55/39.11)
AutoNER	67.00(75.21/60.40)	26.10(43.26/18.69)	67.18(64.63/69.95)	51.39(48.82/54.23)	47.54(43.54/52.35)
LRNT	69.74(79.91/61.87)	23.84(46.94/15.98)	67.69(67.36/68.02)	47.74(46.70/48.83)	46.21(45.60/46.84)
Other Baseline (Reported Results)					
KALM †	76.00(– / –)	–	–	–	–
ConNET ◇	75.57(84.11/68.61)	–	–	–	–
Our BOND **Framework**					
Stage I	75.61(83.76/68.90)	46.61(53.11/41.52)	68.11(66.71/69.56)	59.11(60.14/58.11)	51.55(49.17/54.50)
BOND	81.48(82.05/80.92)	48.01(53.16/43.76)	68.35(67.14/69.61)	65.74(67.37/64.19)	60.07(53.44/68.58)

Experiment: Main Result

Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Tweet	OntoNote5.0	Webpage	Wikigold
Entity Types	4	10	18	4	4
KB Matching	71.40(81.13/63.75)	35.83(40.34/32.22)	59.51(63.86/55.71)	52.45(62.59/45.14)	47.76(47.90/47.63)
Fully-Supervised (Our implementation)					
RoBERTa	90.11(89.14/91.10)	52.19(51.76/52.63)	86.20(84.59/87.88)	72.39(66.29/79.73)	86.43(85.33/87.56)
BiLSTM-CRF	91.21(91.35/91.06)	52.18(60.01/46.16)	86.17(85.99/86.36)	52.34(50.07/54.76)	54.90(55.40/54.30)
Baseline (Our implementation)					
BiLSTM-CRF	59.50(75.50/49.10)	21.77(46.91/14.18)	66.41(68.44/64.50)	43.34(58.05/34.59)	42.92(47.55/39.11)
AutoNER	67.00(75.21/60.40)	26.10(43.26/18.69)	67.18(64.63/69.95)	51.39(48.82/54.23)	47.54(43.54/52.35)
LRNT	69.74(79.91/61.87)	23.84(46.94/15.98)	67.69(67.36/68.02)	47.74(46.70/48.83)	46.21(45.60/46.84)
Other Baseline (Reported Results)					
KALM †	76.00(– / –)	–	–	–	–
ConNET ◇	75.57(84.11/68.61)	–	–	–	–
Our BOND Framework					
Stage I	75.61(83.76/68.90)	46.61(53.11/41.52)	68.11(66.71/69.56)	59.11(60.14/58.11)	51.55(49.17/54.50)
BOND	81.48(82.05/80.92)	48.01(53.16/43.76)	68.35(67.14/69.61)	65.74(67.37/64.19)	60.07(53.44/68.58)

Experiment: Main Result

Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Tweet	OntoNote5.0	Webpage	Wikigold
Entity Types	4	10	18	4	4
KB Matching	71.40(81.13/63.75)	35.83(40.34/32.22)	59.51(63.86/55.71)	52.45(62.59/45.14)	47.76(47.90/47.63)
Fully-Supervised (Our implementation)					
RoBERTa	90.11(89.14/91.10)	52.19(51.76/52.63)	86.20(84.59/87.88)	72.39(66.29/79.73)	86.43(85.33/87.56)
BiLSTM-CRF	91.21(91.35/91.06)	52.18(60.01/46.16)	86.17(85.99/86.36)	52.34(50.07/54.76)	54.90(55.40/54.30)
Baseline (Our implementation)					
BiLSTM-CRF	59.50(75.50/49.10)	21.77(46.91/14.18)	66.41(68.44/64.50)	43.34(58.05/34.59)	42.92(47.55/39.11)
AutoNER	67.00(75.21/60.40)	26.10(43.26/18.69)	67.18(64.63/69.95)	51.39(48.82/54.23)	47.54(43.54/52.35)
LRNT	69.74(79.91/61.87)	23.84(46.94/15.98)	67.69(67.36/68.02)	47.74(46.70/48.83)	46.21(45.60/46.84)
Other Baseline (Reported Results)					
KALM †	76.00(– / –)	–	–	–	–
ConNET ◇	75.57(84.11/68.61)	–	–	–	–
Our BOND Framework					
Stage I	75.61(83.76/68.90)	46.61(53.11/41.52)	68.11(66.71/69.56)	59.11(60.14/58.11)	51.55(49.17/54.50)
BOND	81.48(82.05/80.92)	48.01(53.16/43.76)	68.35(67.14/69.61)	65.74(67.37/64.19)	60.07(53.44/68.58)

Experiment: Ablation

Table 3: Ablation Study: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Wikigold
Stage I		
Stage I	75.61(83.76/68.90)	51.55(49.17/54.50)
Stage I w/o pre-train	36.66(37.49/35.75)	18.31(18.14/18.50)
Stage I w/o early stop	72.11(81.65/64.57)	49.68(48.67/50.74)
Stage I w/ MT	76.30(82.92/70.67)	46.68(49.82/43.91)
Stage I w/ VAT	76.38(82.58/71.04)	47.54(50.02/45.30)
Stage I + Stage II		
BOND †	77.28(83.42/71.98)	56.90(54.32/59.74)
BOND w/ soft	80.18(81.56/78.84)	58.64(58.29/65.79)
BOND w/ soft+high conf	81.48(82.05/80.92)	60.07(53.44/68.58)
BOND w/ reinit	78.17(85.05/72.31)	58.55(55.31/62.19)
BOND w/ soft+reinit	76.92(83.39/71.38)	54.09(50.72/57.94)
BOND w/ MT	77.16(82.79/72.25)	57.93(55.66/60.39)
BOND w/ VAT	77.64(85.62/70.69)	57.39(55.05/59.41)

Experiment: Ablation

Table 3: Ablation Study: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Wikigold
Stage I		
Stage I	75.61(83.76/68.90)	51.55(49.17/54.50)
Stage I w/o pre-train	36.66(37.49/35.75)	18.31(18.14/18.50)
Stage I w/o early stop	72.11(81.65/64.57)	49.68(48.67/50.74)
Stage I w/ MT	76.30(82.92/70.67)	46.68(49.82/43.91)
Stage I w/ VAT	76.38(82.58/71.04)	47.54(50.02/45.30)
Stage I + Stage II		
BOND †	77.28(83.42/71.98)	56.90(54.32/59.74)
BOND w/ soft	80.18(81.56/78.84)	58.64(58.29/65.79)
BOND w/ soft+high conf	81.48(82.05/80.92)	60.07(53.44/68.58)
BOND w/ reinit	78.17(85.05/72.31)	58.55(55.31/62.19)
BOND w/ soft+reinit	76.92(83.39/71.38)	54.09(50.72/57.94)
BOND w/ MT	77.16(82.79/72.25)	57.93(55.66/60.39)
BOND w/ VAT	77.64(85.62/70.69)	57.39(55.05/59.41)

Experiment: Ablation

Table 3: Ablation Study: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Wikigold
Stage I		
Stage I	75.61(83.76/68.90)	51.55(49.17/54.50)
Stage I w/o pre-train	36.66(37.49/35.75)	18.31(18.14/18.50)
Stage I w/o early stop	72.11(81.65/64.57)	49.68(48.67/50.74)
Stage I w/ MT	76.30(82.92/70.67)	46.68(49.82/43.91)
Stage I w/ VAT	76.38(82.58/71.04)	47.54(50.02/45.30)
Stage I + Stage II		
BOND †	77.28(83.42/71.98)	56.90(54.32/59.74)
BOND w/ soft	80.18(81.56/78.84)	58.64(58.29/65.79)
BOND w/ soft+high conf	81.48(82.05/80.92)	60.07(53.44/68.58)
BOND w/ reinit	78.17(85.05/72.31)	58.55(55.31/62.19)
BOND w/ soft+reinit	76.92(83.39/71.38)	54.09(50.72/57.94)
BOND w/ MT	77.16(82.79/72.25)	57.93(55.66/60.39)
BOND w/ VAT	77.64(85.62/70.69)	57.39(55.05/59.41)

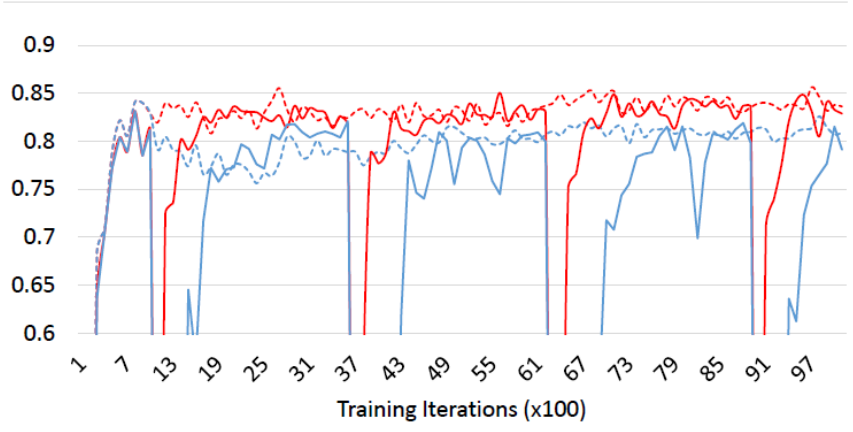
Experiment: Ablation

Table 3: Ablation Study: F_1 Score (Precision/Recall) (in %)

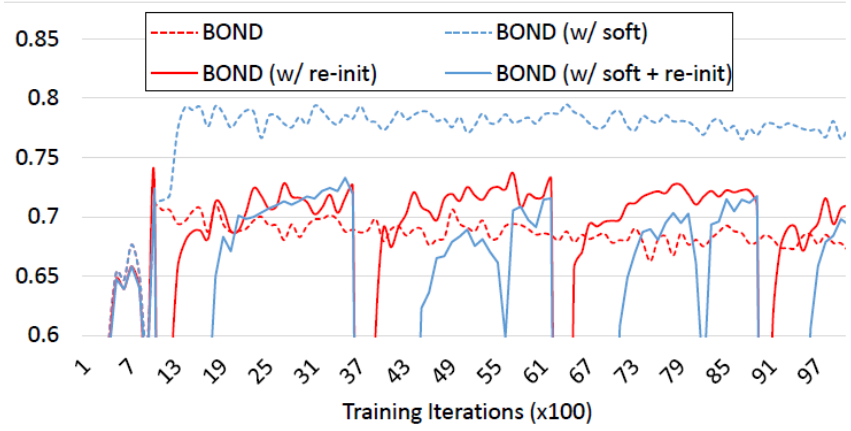
Method	CoNLL03	Wikigold
Stage I		
Stage I	75.61(83.76/68.90)	51.55(49.17/54.50)
Stage I w/o pre-train	36.66(37.49/35.75)	18.31(18.14/18.50)
Stage I w/o early stop	72.11(81.65/64.57)	49.68(48.67/50.74)
Stage I w/ MT	76.30(82.92/70.67)	46.68(49.82/43.91)
Stage I w/ VAT	76.38(82.58/71.04)	47.54(50.02/45.30)
Stage I + Stage II		
BOND †	77.28(83.42/71.98)	56.90(54.32/59.74)
BOND w/ soft	80.18(81.56/78.84)	58.64(58.29/65.79)
BOND w/ soft+high conf	81.48(82.05/80.92)	60.07(53.44/68.58)
BOND w/ reinit	78.17(85.05/72.31)	58.55(55.31/62.19)
BOND w/ soft+reinit	76.92(83.39/71.38)	54.09(50.72/57.94)
BOND w/ MT	77.16(82.79/72.25)	57.93(55.66/60.39)
BOND w/ VAT	77.64(85.62/70.69)	57.39(55.05/59.41)

Experiment: Ablation

(a) F_1 score



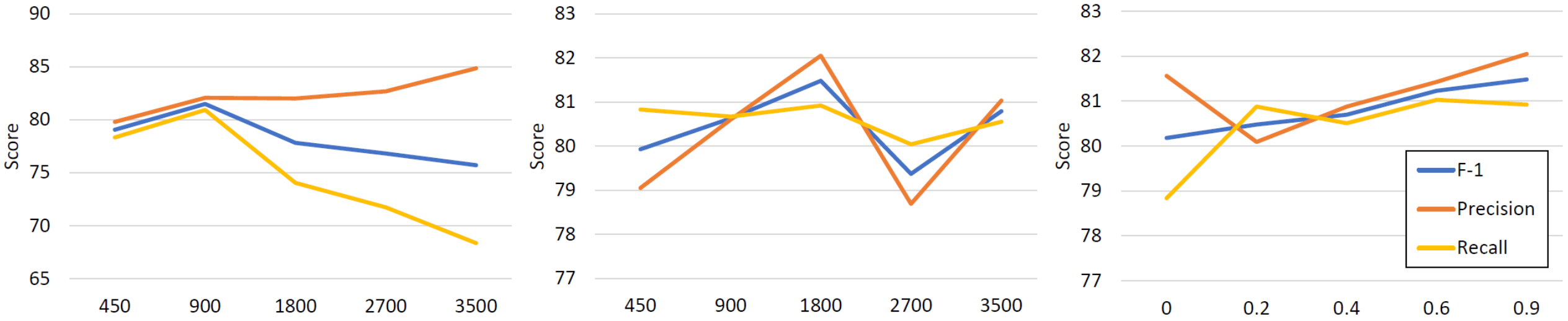
(b) Precision



(c) Recall

Figure 6: Learning Curves of BOND, BOND **(w/ reinit),** BOND **(w/ soft) and** BOND **(w/ soft + reinit)**

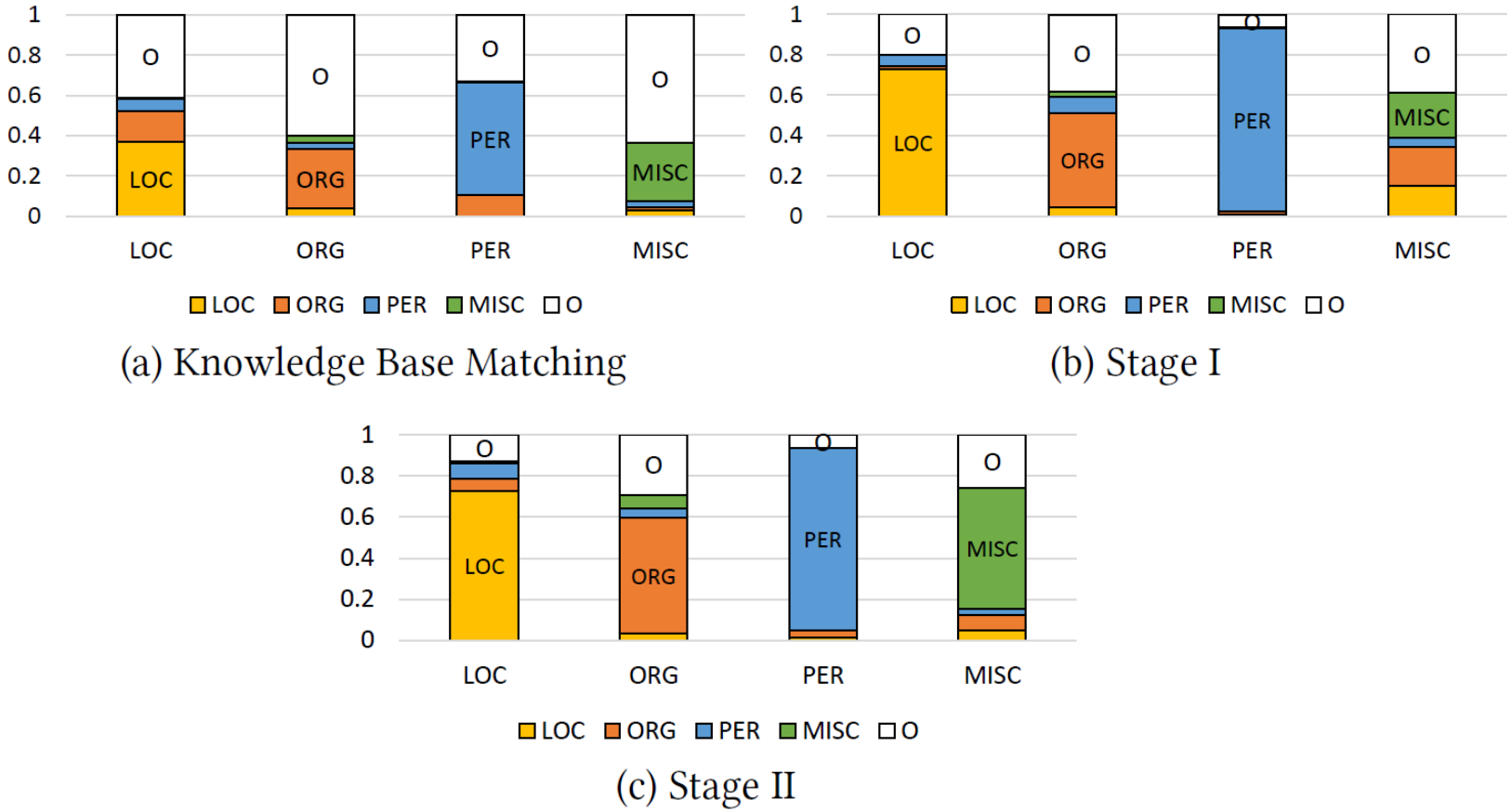
Experiment: Parameter Study



(a) The Early Stopping Time of Stage I – T_1 (b) The Early Stopping Time in Stage II – T_3 (c) The Confidence Threshold of Stage II – ϵ

Figure 7: Parameter Study using CoNLL03: F_1, Precision, Recall on Testing Set (in %)

Experiment: Error Analysis



(a) Knowledge Base Matching

(b) Stage I

(c) Stage II

Figure 8: Recall of Knowledge Base Matching and different stages of BOND.
The horizontal axis denotes the true entity type.
The segments in a bar denote the portions of the entities being classified into different entity types.

Thank You!
To Find Out More?

- Arxiv: https://arxiv.org/abs/2006.15509
- Git: https://github.com/cliang1453/BOND