

# Knowledge Base Population from Text and Graphs

Lucas Sterckx

# About Me

- Fourth-year PhD Student
- Research Topics:
  - Relation Extraction
  - Keyphrase Extraction
  - Sequence-to-Sequence Models
- Currently visiting Machine Intelligence Lab under supervision of **prof. Bill Byrne** and **dr. Jason Naradowsky**



# NLP @ Ghent University

- Part of Internet Technology and Data Science Lab
- Initial focus: Information Retrieval
- NLP and Information Extraction since 2013
- Projects with Flanders' major media providers
  - Named Entity Recognition
  - Named Entity Linking
  - Text Classification
  - Keyphrase extraction



# Outline

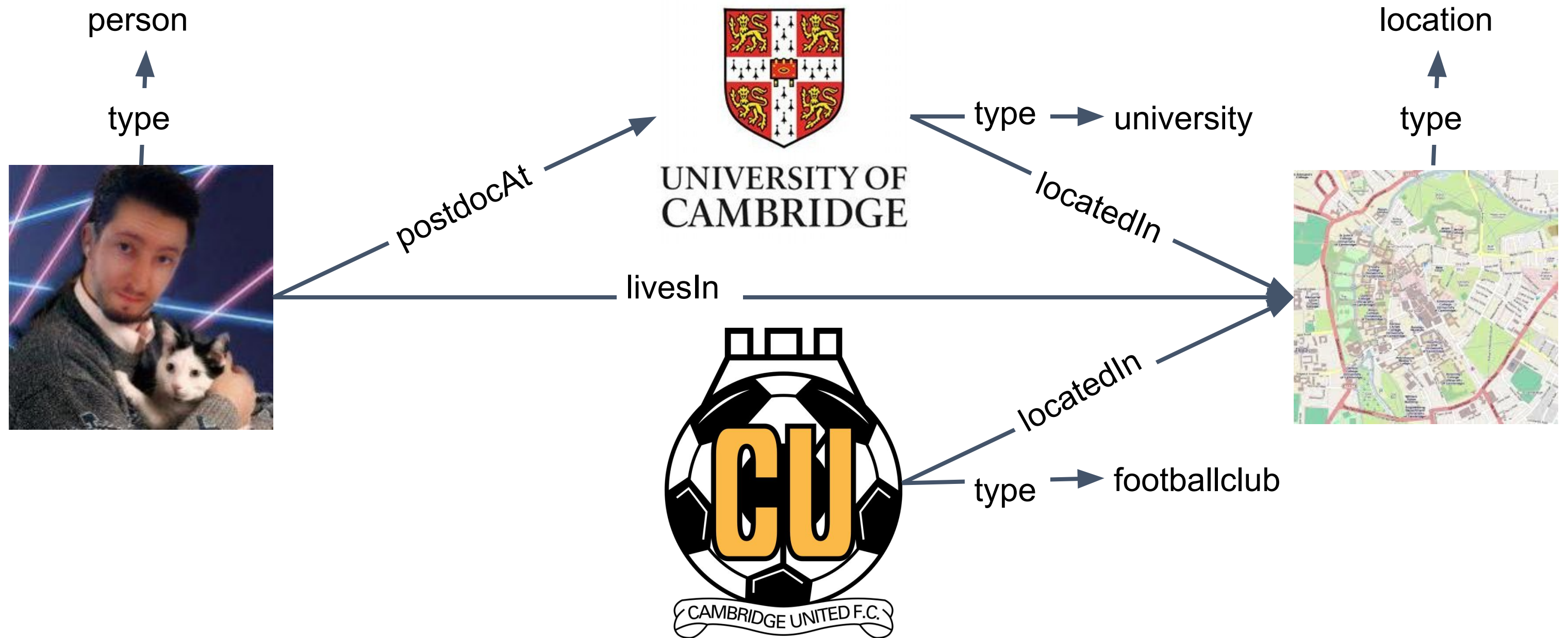
## 1. Knowledge Bases

## 2. Knowledge Base Population

- a. Knowledge Extraction from Text - TAC KBP
- b. Link Prediction

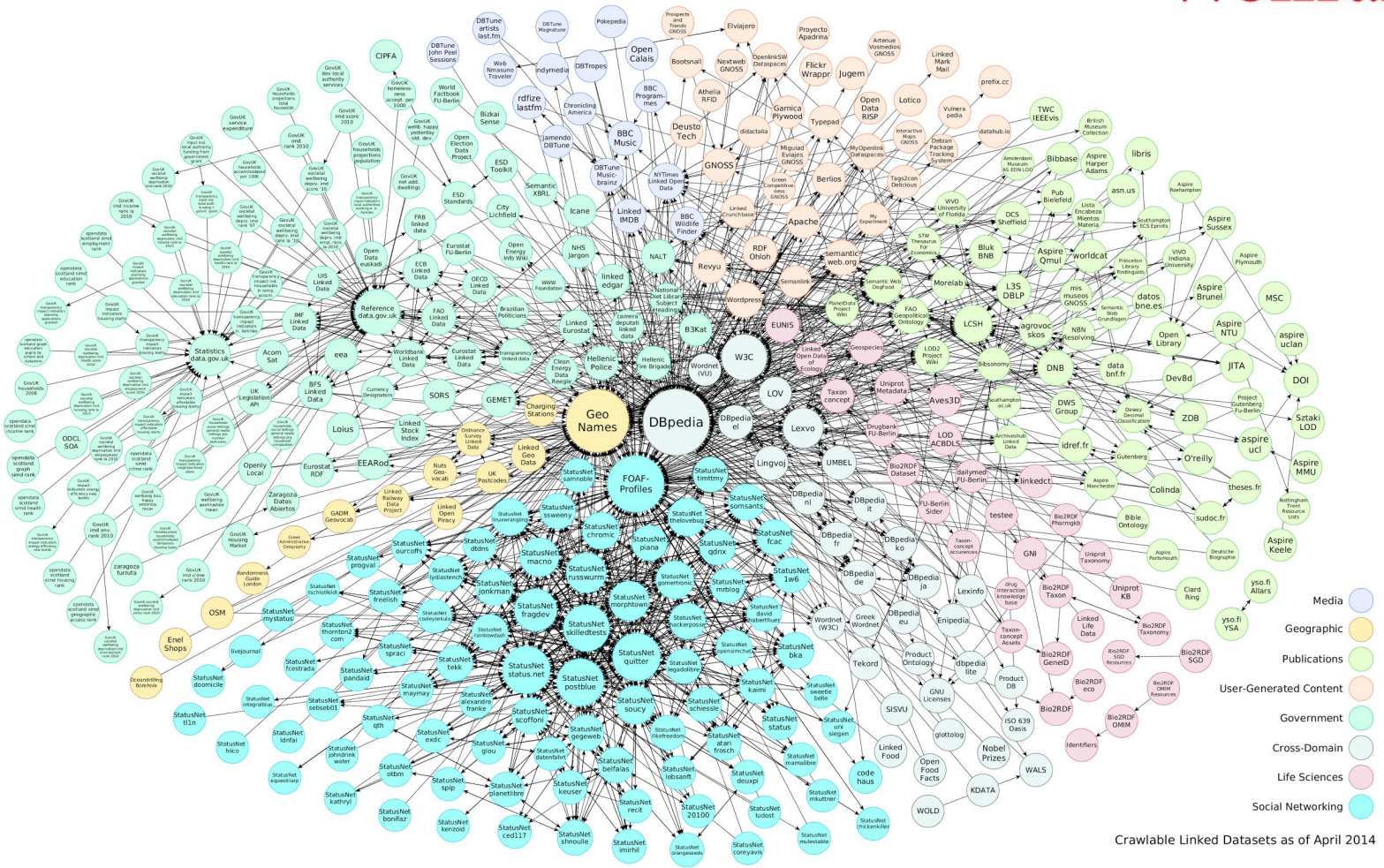
## 3. Research Topic at Cambridge

# Knowledge Bases as Labeled Graphs



Comprehensive and semantically organized **machine-readable** collection of universally relevant or domain-specific **entities**, **classes**, and **SPO facts** (attributes, relations)

# Modern Knowledge Graphs



# Applications

cambridge uk

Alle Afbeeldingen Maps Nieuws Video's Meer Instellingen Tools

Ongeveer 298.000.000 resultaten (0,96 seconden)

## Cambridge - Wikipedia

<https://en.wikipedia.org/wiki/Cambridge> ▼ Vertaal deze pagina

Cambridge is a university city and the county town of Cambridgeshire, England, on the River Cam about 50 miles (80 km) north of London. At the United ...

University of Cambridge · Cambridgeshire · Cambridge, Massachusetts · St Ives

## University of Cambridge

<https://www.cam.ac.uk/> ▼ Vertaal deze pagina

The mission of the University of Cambridge is to contribute to society through the pursuit of education, learning and research at the highest international levels of ...

## Cambridge Hotels, Things to Do, Events - Official Cambridge Tourist ...

[www.visitcambridge.org/](http://www.visitcambridge.org/) ▼ Vertaal deze pagina

Official Visitor Information for Cambridge, England. Find things to do, hotels and accommodation, attractions, events, restaurants, shopping maps – everything ...

## Cambridge, United Kingdom - TripAdvisor

<https://www.tripadvisor.com> > ... > England > Cambridgeshire ▼ Vertaal deze pagina

Cambridge Tourism: TripAdvisor has 128820 reviews of Cambridge Hotels, Attractions, and Restaurants making it your best Cambridge resource.

## Cambridge 2017: Best of Cambridge,

<https://www.tripadvisor.co.uk> > ... > England > Can

Cambridge Tourism: TripAdvisor has 128820 review Restaurants making it your best Cambridge resourc

## Cambridge - Lonely Planet

<https://www.lonelyplanet.com/england/eastern-er> ... and tradition and renowned for its quirky rituals, Ci minds England's two most venerable university citie

## Cambridge City Council

<https://www.cambridge.gov.uk/> ▼ Vertaal deze pa

Local and community information with sections on er

## Cambridge Tourist Information

[www.cambridgetouristinformation.co.uk/](http://www.cambridgetouristinformation.co.uk/) ▼ Verta

Kyan is a Cambridge born musician and he's going http://www.kings.cam.ac.uk/events/chapel-services.f

## Cambridge travel guide - Wikitravel

[wikitravel.org/en/Cambridge\\_\(England\)](http://wikitravel.org/en/Cambridge_(England)) ▼ Vertaa

Cambridge [1] is a university city in Cambridgeshir on the Backs, of green open spaces and cattle graziz

bezienswaardigheden cambridge omgeving

Alle Maps Afbeeldingen Nieuws Shopping Meer Instellingen Tools

## Cambridge > Bezienswaardigheden

Fitzwilliam Museum Museum en architectuur	Anglesey Abbey Tuin en hortus botanicus	Sedgwick Museum of Earth Sci...	Kathedraal van Ely Kathedraal, architectuur en geschiedenis	Cambridge Museum of Technolo...
Cambridge University Botanic G... Hortus botanicus en tuin	Church of St Mary the Great, Ca... Kerkgebouw	Museum of Archaeology and Ant... Museum en geschiedenis	Whipple Museum of the History ...	Cambridge University Museum o... Museum
The Backs Tuin, park en rivier	Heilig Grafkerk Kerkgebouw en architectuur	Museum of Cambridge Museum	Parker's Piece	Museum of Classical Archaeolog... Museum

## Cambridge: attracties en bezienswaardigheden. - Weekendplanner

[www.weekendplanner.nl](http://www.weekendplanner.nl) > Engeland > Cambridge ▼

Cambridge is een mooie stad vlakbij Londen. ... De attracties en bezienswaardigheden in Cambridge ... Informatie voor een bezoek aan Londen en omgeving ...

## Wat te doen in Cambridge: de 10 beste activiteiten - TripAdvisor

<https://www.tripadvisor.nl> > ... > Verenigd Koninkrijk > Cambridgeshire > Cambridge ▼

cambridge population

Alle Afbeeldingen Maps Nieuws Shopping Meer Instellingen Tools

Ongeveer 222.000.000 resultaten (0,78 seconden)

## Cambridge / Bevolking

123.900

2011



Feedback



# Applications



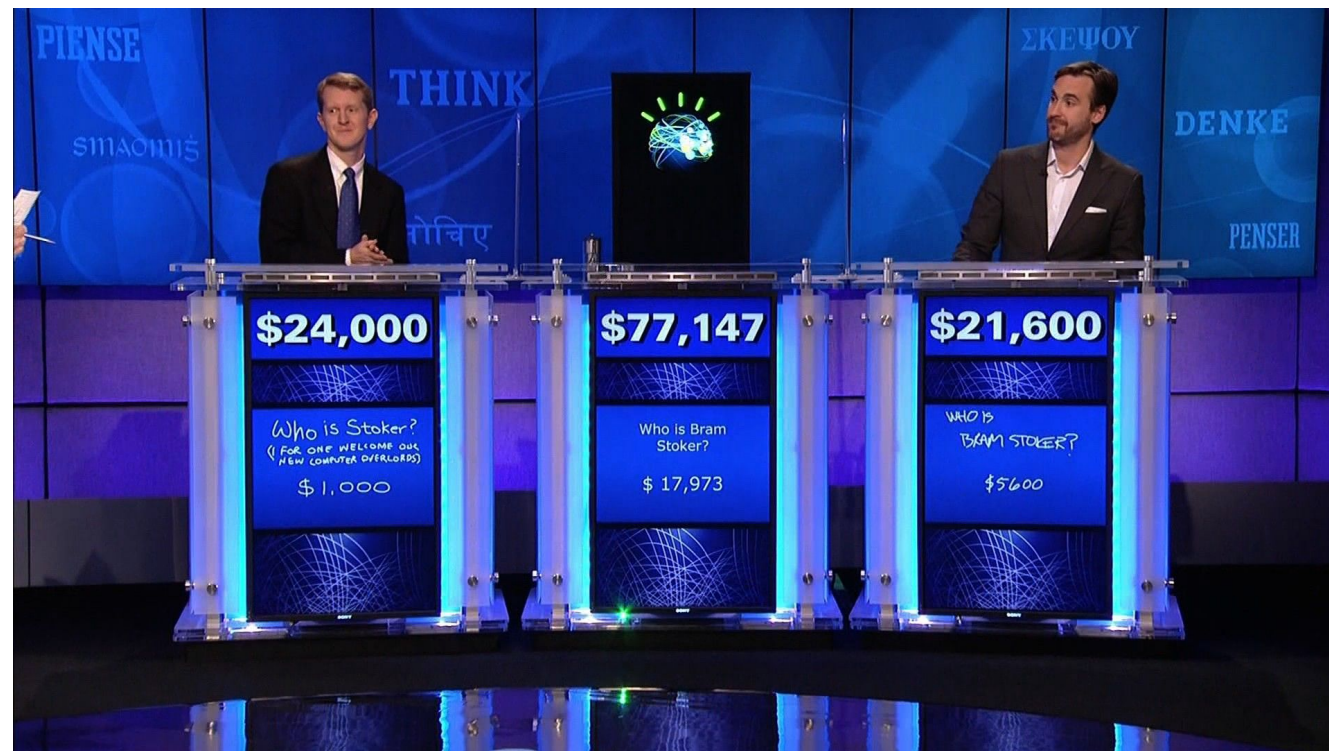
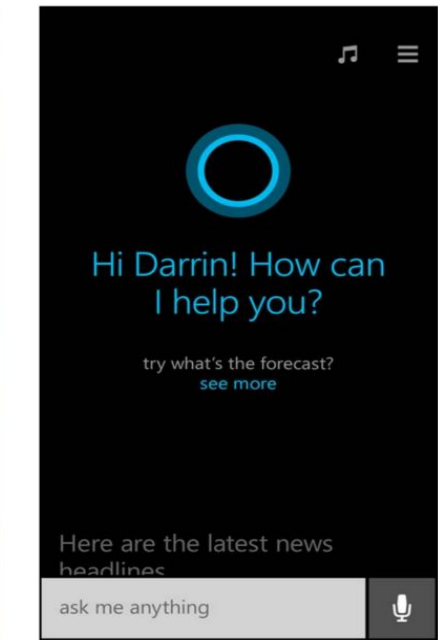
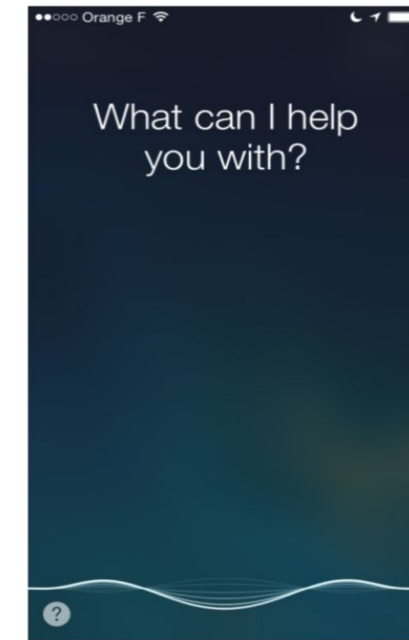
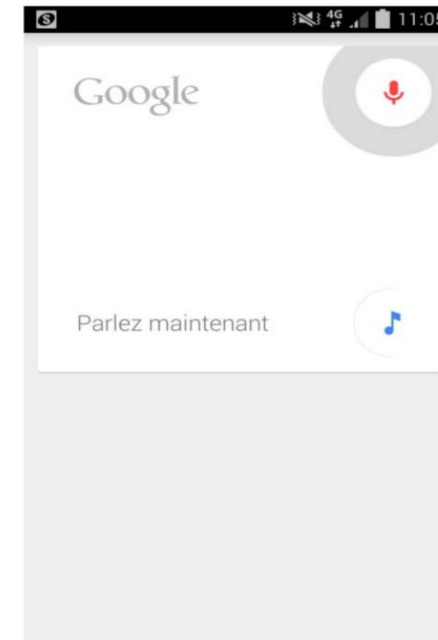
when did the beatles break up

Web Apps Examples Random

Input interpretation:  
The Beatles (music act) end date

Open code

Result:  
Friday, April 10, 1970





# Knowledge Bases: Research Challenges

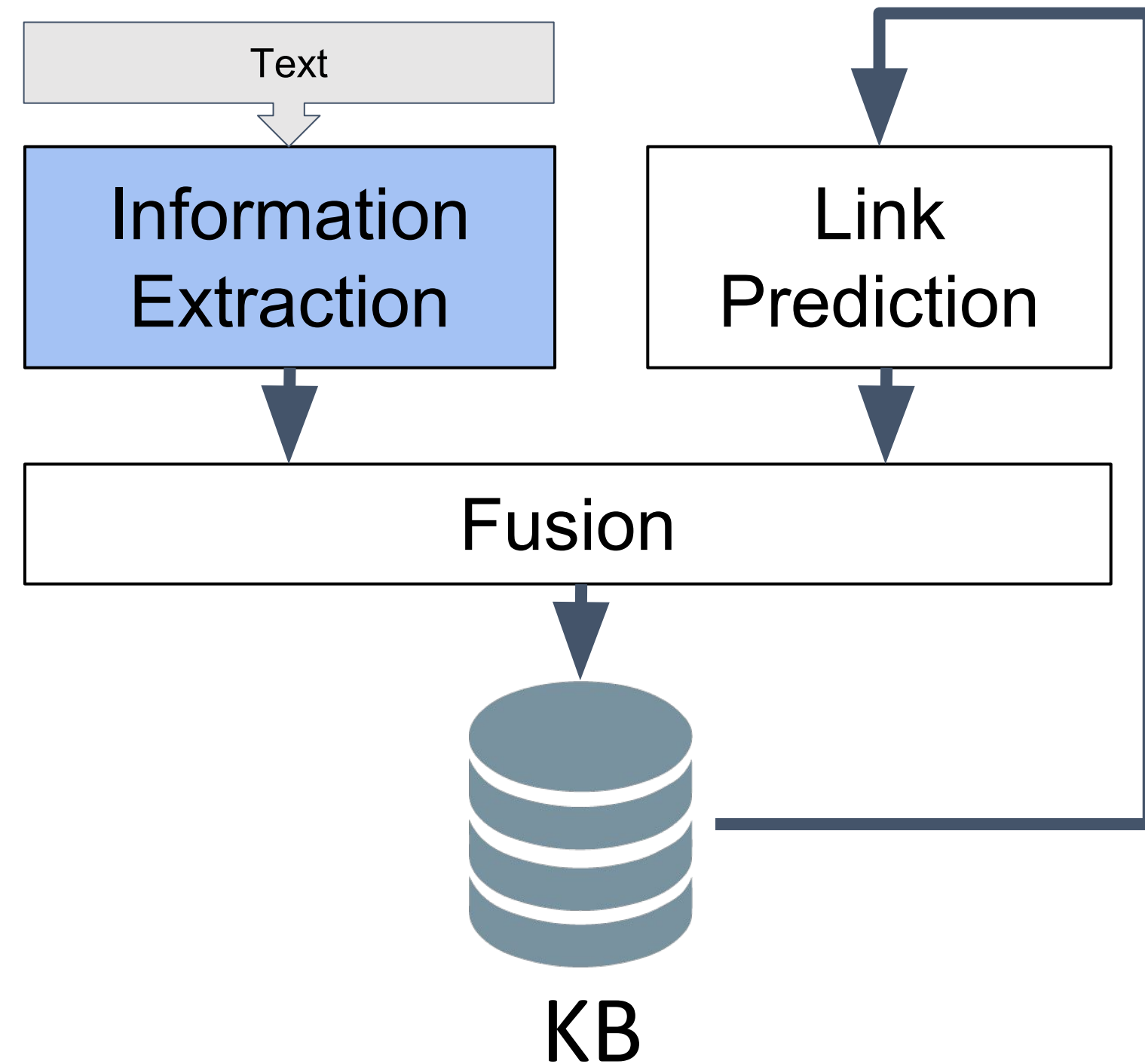
- **Population: Knowledge bases are incomplete**
  - Knowledge extraction from text
  - Link prediction
- **Validation: Knowledge bases contain errors**
  - Entity resolution
  - Error detection, trustworthiness
- **Interface: How to easily access knowledge**
  - Semantic parsing
  - Question answering
- **General AI: can AI emerge from Knowledge Graphs?**
  - Automatic reasoning and planning
  - Generalization and abstraction

# Knowledge Bases: Research Challenges

- **Population: Knowledge bases are incomplete**
  - Knowledge extraction from text
  - Link prediction
- **Validation: Knowledge bases contain errors**
  - Entity resolution
  - Error detection, trustworthiness
- **Interface: How to easily access knowledge**
  - Semantic parsing
  - Question answering
- **General AI: can AI emerge from Knowledge Graphs?**
  - Automatic reasoning and planning
  - Generalization and abstraction

# Outline

1. Knowledge Bases
- 2. Knowledge Base Population**
  - a. Knowledge Extraction from Text
  - b. Link Prediction
3. Research Topic at Cambridge



# Text Analysis Conference (TAC)



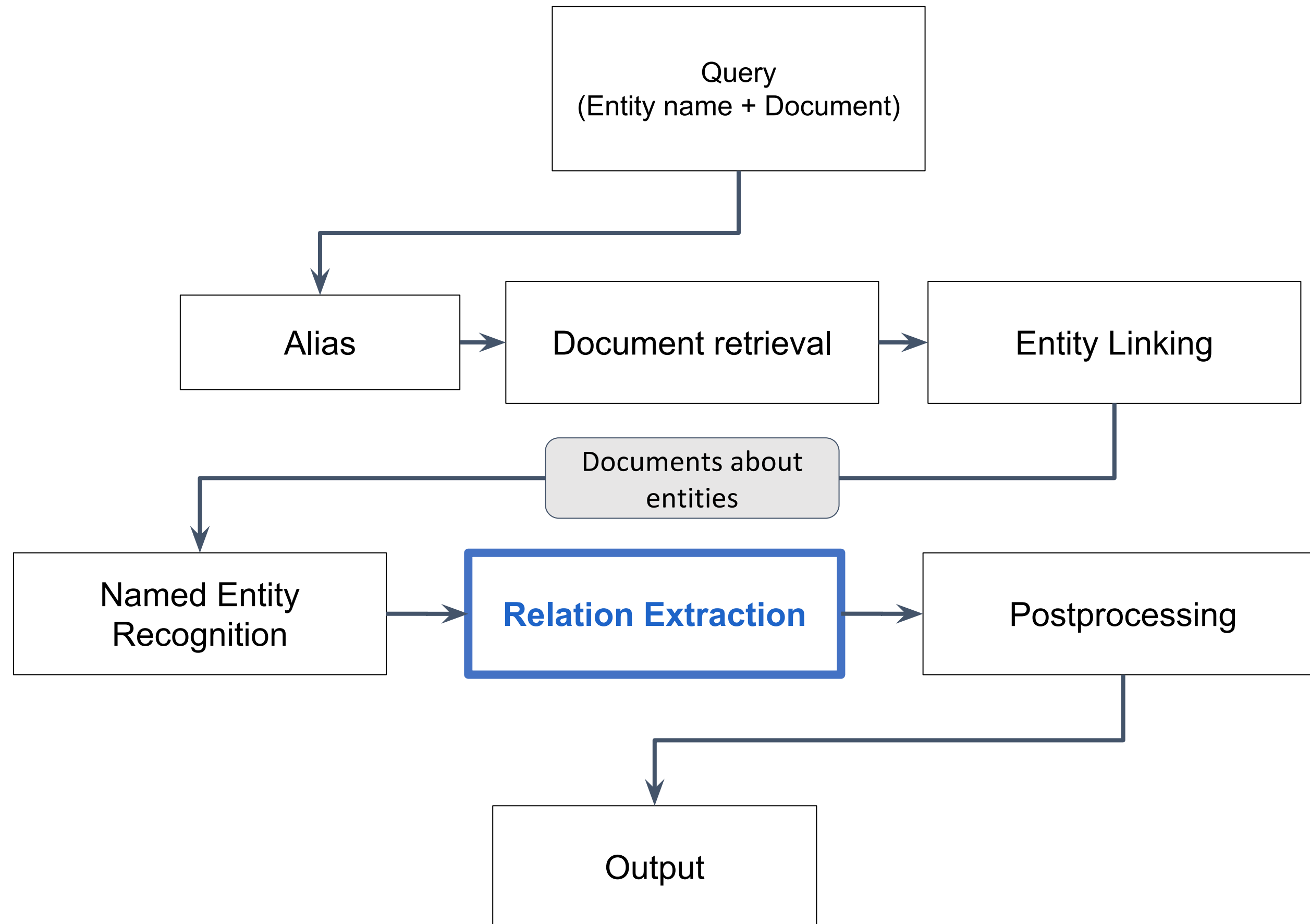
- **Benchmark to promote research** on automated systems that discover information about named entities and incorporate this information in a knowledge source, or database
- Continuation of previous conferences and evaluations, such as MUC and ACE
- Organised by **NIST**, sponsored by US department of Defense
- Extract **41 relations** for **50 persons** and **50 organizations** from **4 million documents**

```
<query id="SF_002">  
  <name>PhillyInquirer</name>  
  <docid>eng-NG-31-141808-9966244</docid>  
  <beg>757</beg>  
  <end>770</end>  
  <enttype>ORG</enttype>  
</query>
```

**Person Slots**

<b>Name</b>	<b>Type</b>	<b>List?</b>
per:alternate_names	Name	Yes
per:date_of_birth	Value	
per:age	Value	
per:country_of_birth	Name	
per:stateorprovince_of_birth	Name	
per:city_of_birth	Name	

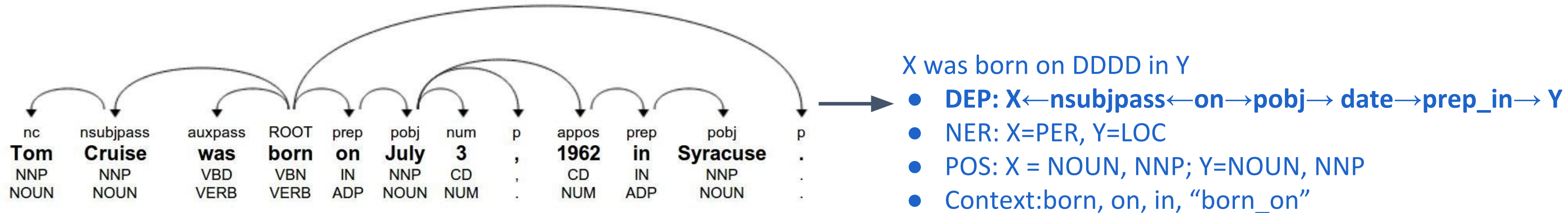
# Ghent University at TAC KBP



# Supervised Relation Extraction

- Extracting semantic relations between sets of grounded entities
- Train classifiers from +/- Examples

< person, city\_of\_birth, location >



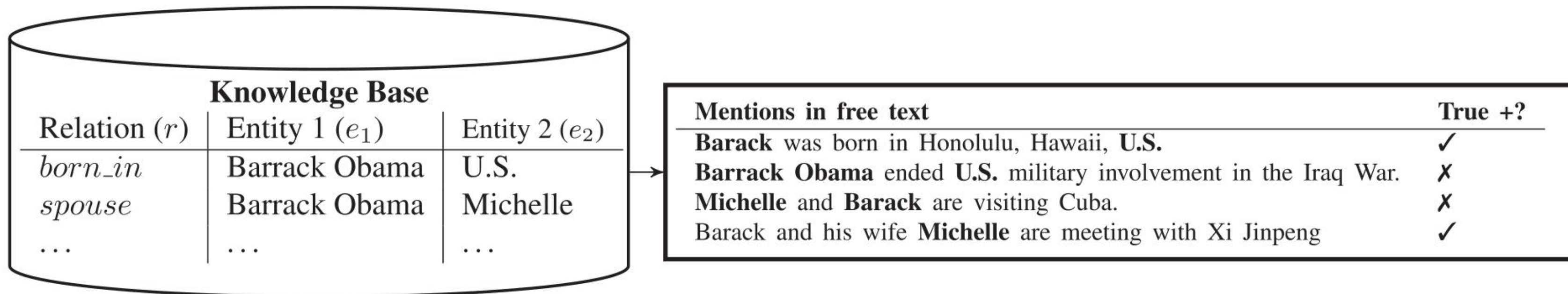
## ✓ Pro's

- High quality training data
- Explicit negative examples

## - Con's

- Expensive!
- Can't generalize to other relations and domains

# Distant Supervision



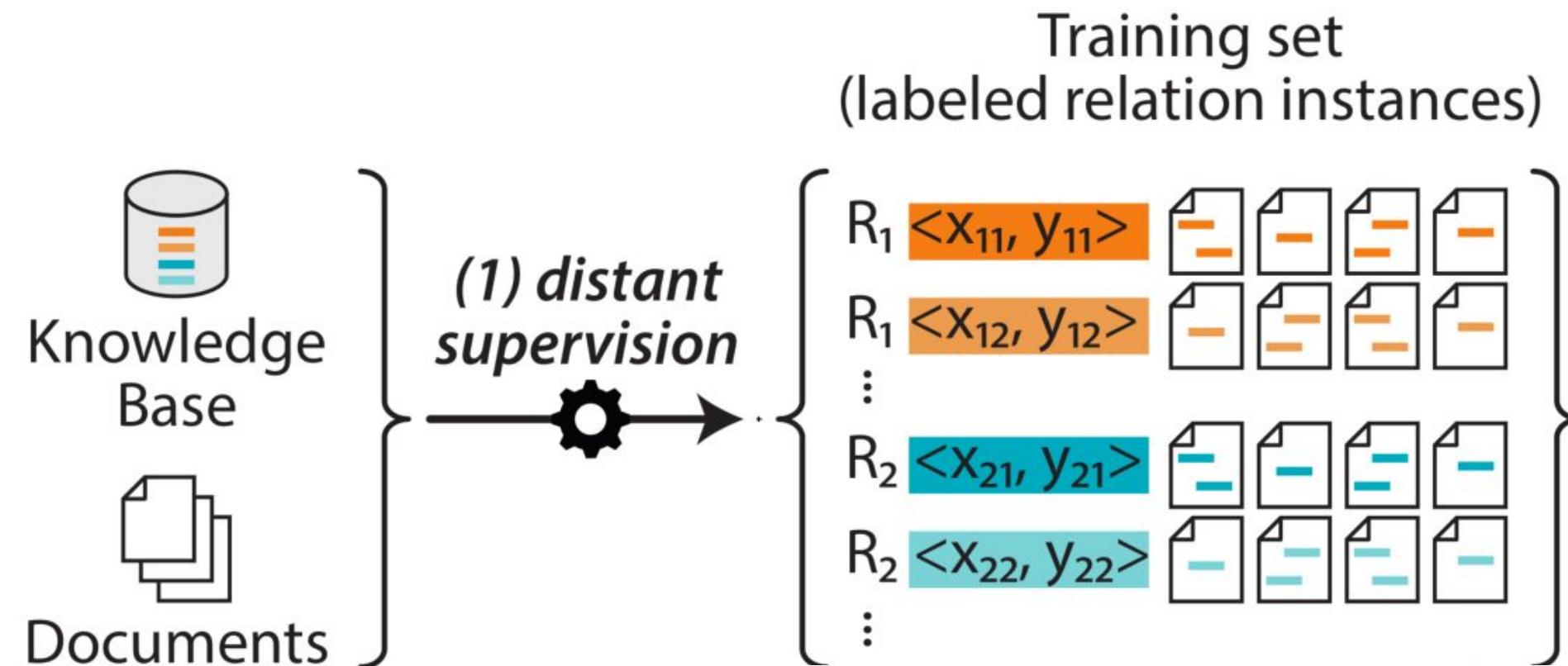
- Existing **Knowledge base + Unlabeled text**

1. Collect many pairs of entities **co-occurring** in sentences from the corpus (Mintz, 2009)

- Noise

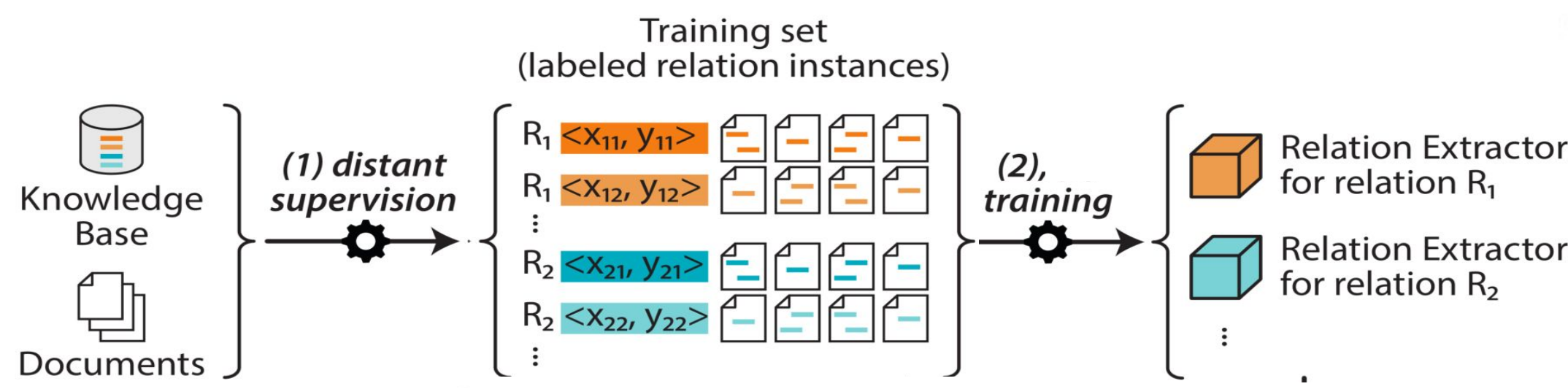
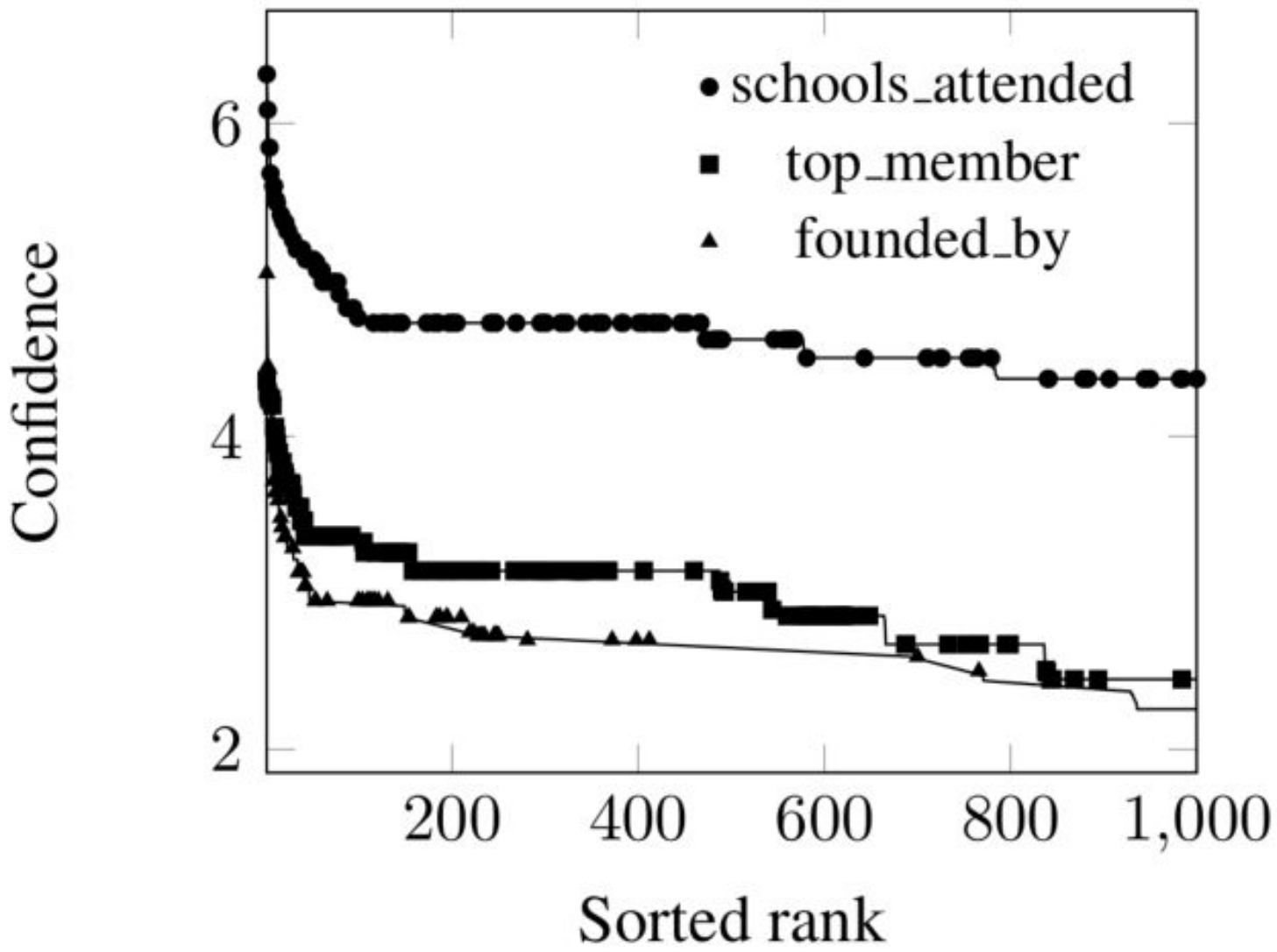
# Guiding Bootstrapped Relation Extractors

Knowledge Base Population using Semantic Label Propagation  
(Knowledge Based Systems, Sterckx, 2016)



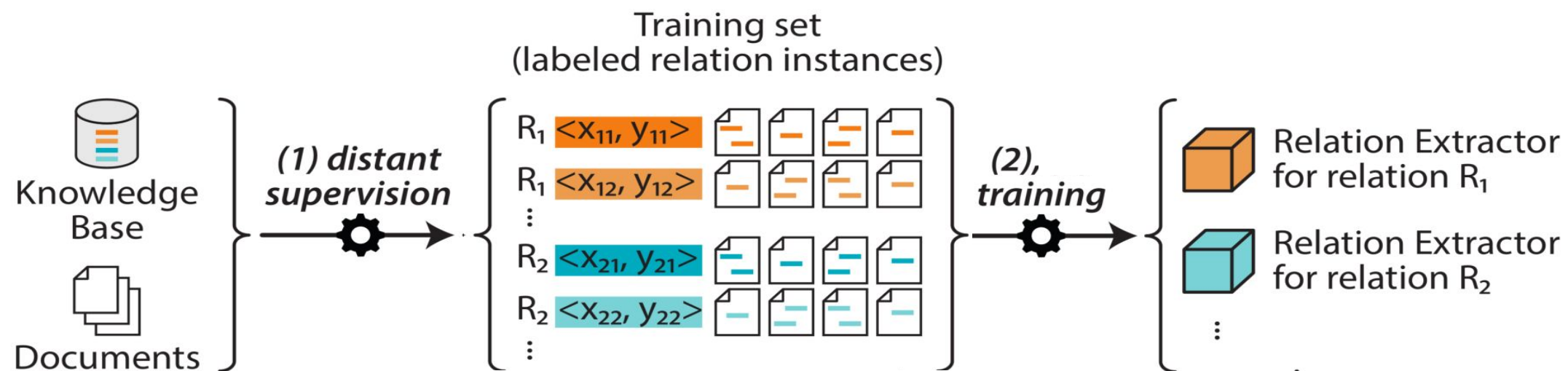


# Guiding Bootstrapped Relation Extractors



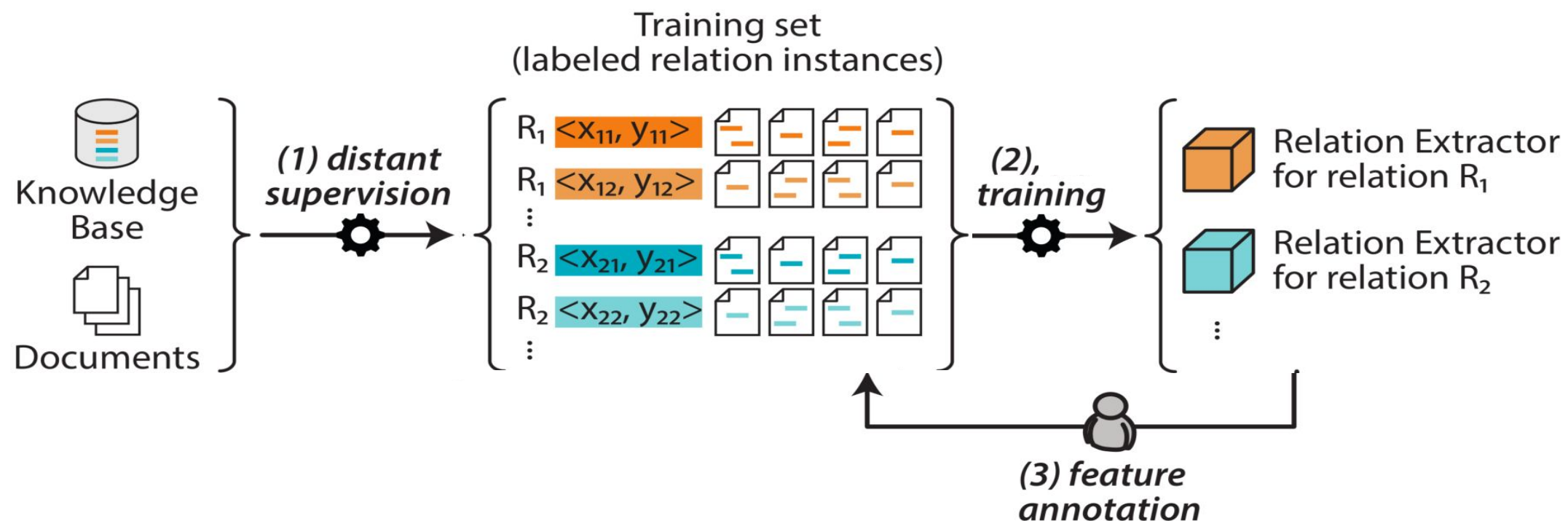
# Guiding Bootstrapped Relation Extractors

Relation	Top SDP
top_members_employees	PER $\xleftarrow{\text{appos}}$ executive $\xrightarrow{\text{prep\_of}}$ ORG
	PER $\xleftarrow{\text{appos}}$ chairman $\xrightarrow{\text{appos}}$ ORG
	ORG $\xleftarrow{\text{nn}}$ founder $\xrightarrow{\text{prep\_of}}$ PER
children	PER-2 $\xleftarrow{\text{appos}}$ son $\xrightarrow{\text{prep\_of}}$ PER-1
	PER-1 $\xleftarrow{\text{appos}}$ father $\xrightarrow{\text{prep\_of}}$ PER-2
	PER-2 $\xleftarrow{\text{nn}}$ grandson $\xrightarrow{\text{prep\_of}}$ PER-1
city_of_birth	PER $\xleftarrow{\text{rcmod}}$ born $\xrightarrow{\text{prep\_in}}$ LOC
	PER $\xleftarrow{\text{nsubj}}$ mayor $\xrightarrow{\text{prep\_of}}$ LOC
	PER $\xleftarrow{\text{appos}}$ historian $\xrightarrow{\text{prep\_from}}$ LOC



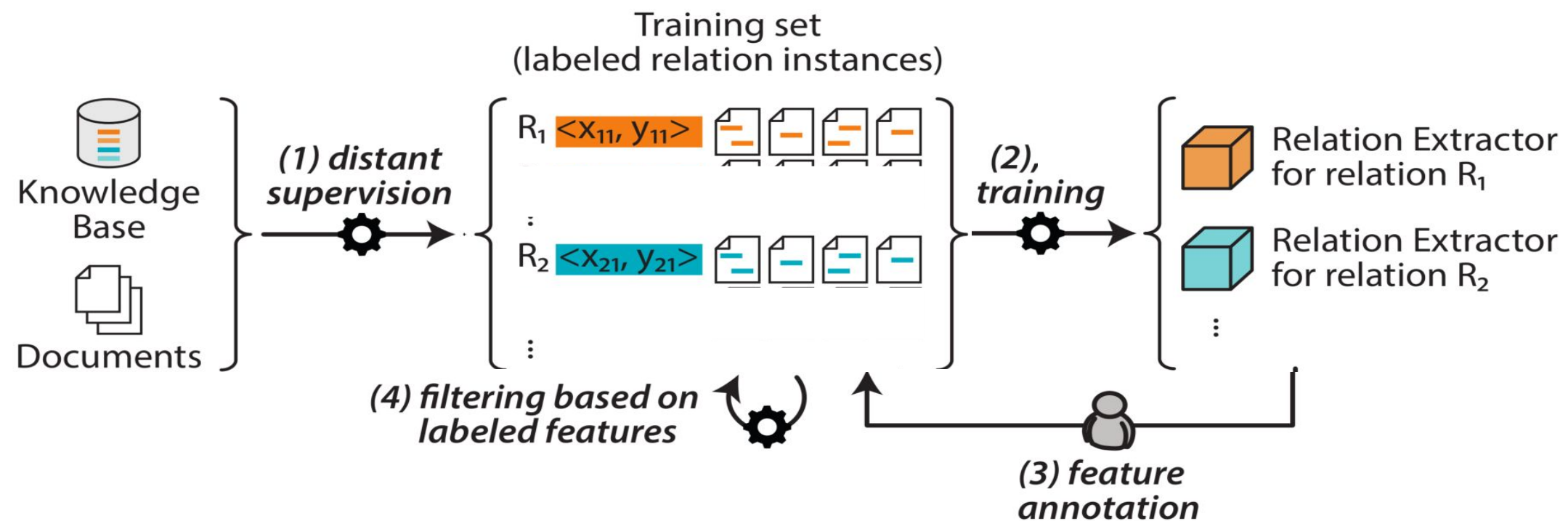
# Guiding Bootstrapped Relation Extractors

Relation	Top SDP	Assessment
top_members_employees	PER $\xleftarrow{\text{appos}}$ executive $\xrightarrow{\text{prep\_of}}$ ORG	✓
	PER $\xleftarrow{\text{appos}}$ chairman $\xrightarrow{\text{appos}}$ ORG	✓
	ORG $\xleftarrow{\text{nn}}$ founder $\xrightarrow{\text{prep\_of}}$ PER	✗
children	PER-2 $\xleftarrow{\text{appos}}$ son $\xrightarrow{\text{prep\_of}}$ PER-1	✓
	PER-1 $\xleftarrow{\text{appos}}$ father $\xrightarrow{\text{prep\_of}}$ PER-2	✓
	PER-2 $\xleftarrow{\text{nn}}$ grandson $\xrightarrow{\text{prep\_of}}$ PER-1	✗
city_of_birth	PER $\xleftarrow{\text{rcmod}}$ born $\xrightarrow{\text{prep\_in}}$ LOC	✓
	PER $\xleftarrow{\text{nsubj}}$ mayor $\xrightarrow{\text{prep\_of}}$ LOC	✗
	PER $\xleftarrow{\text{appos}}$ historian $\xrightarrow{\text{prep\_from}}$ LOC	✗

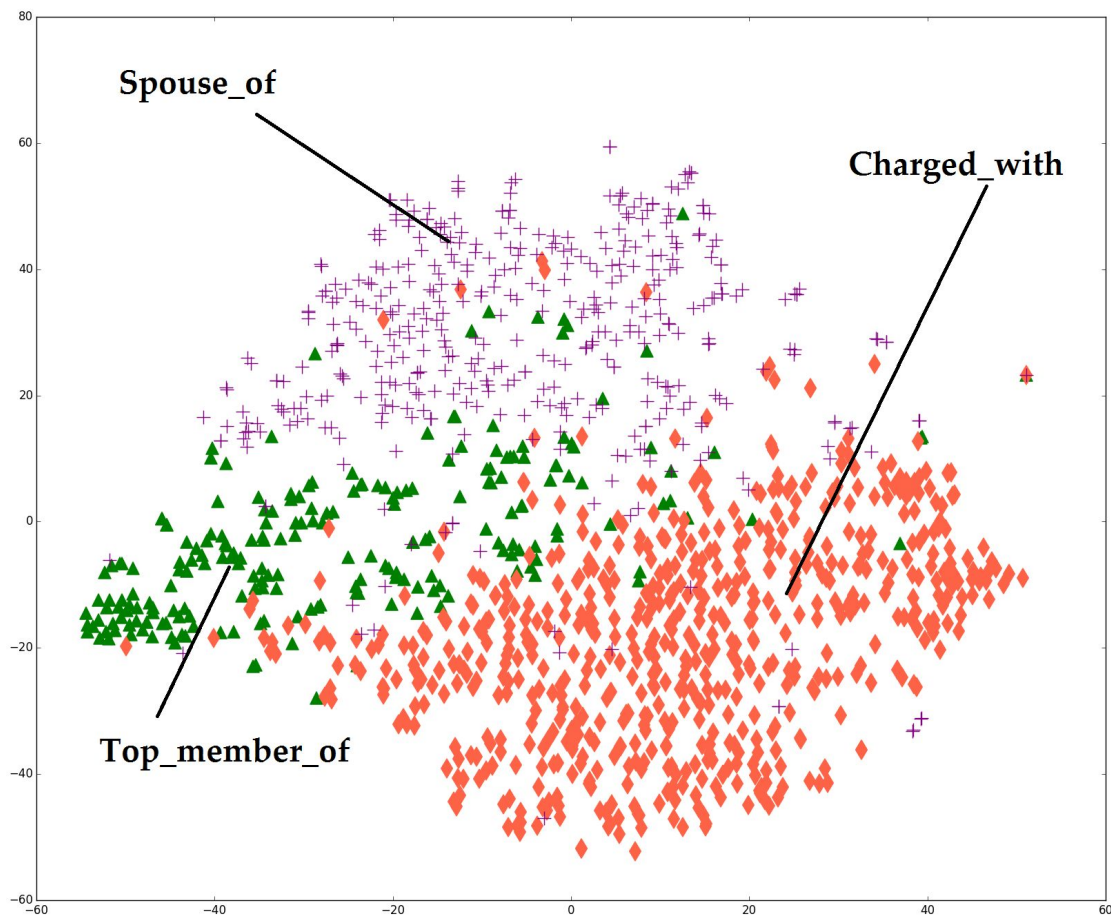


# Guiding Bootstrapped Relation Extractors

- Filter non-labeled patterns
  - Include weaker features in log-linear classifiers
  - Regularize
- Recall ↓↓



# Guiding Bootstrapped Relation Extractors



Labeled Shortest Dependency Path

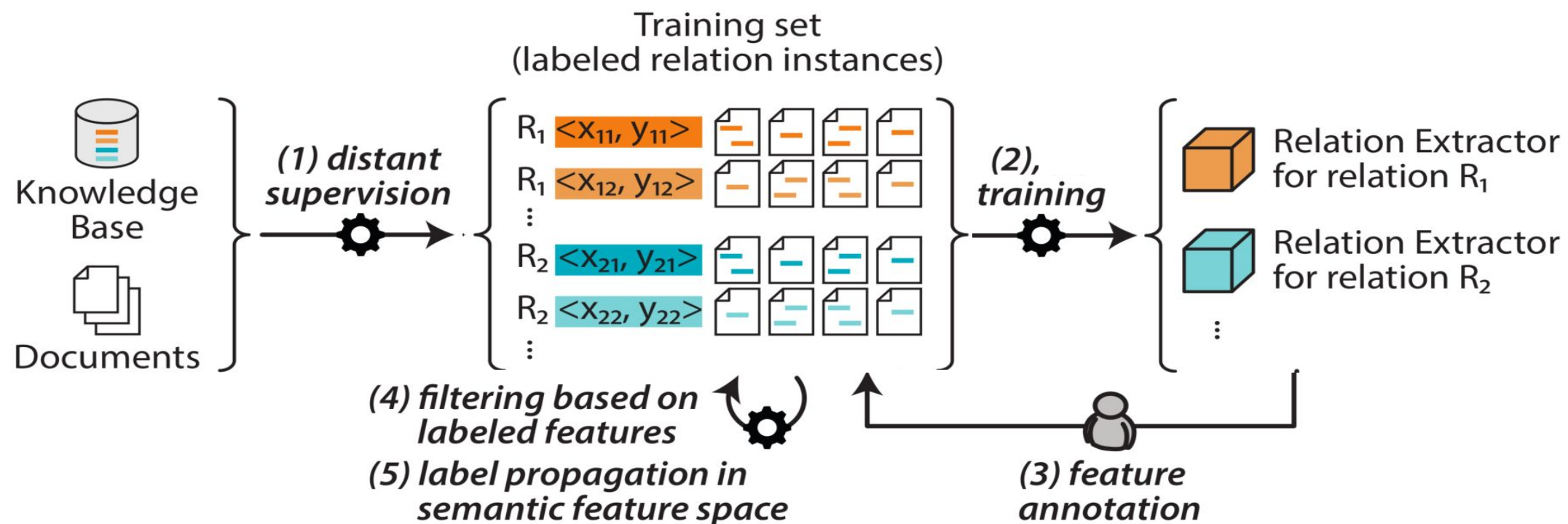
Embedding

Cosine Similarity

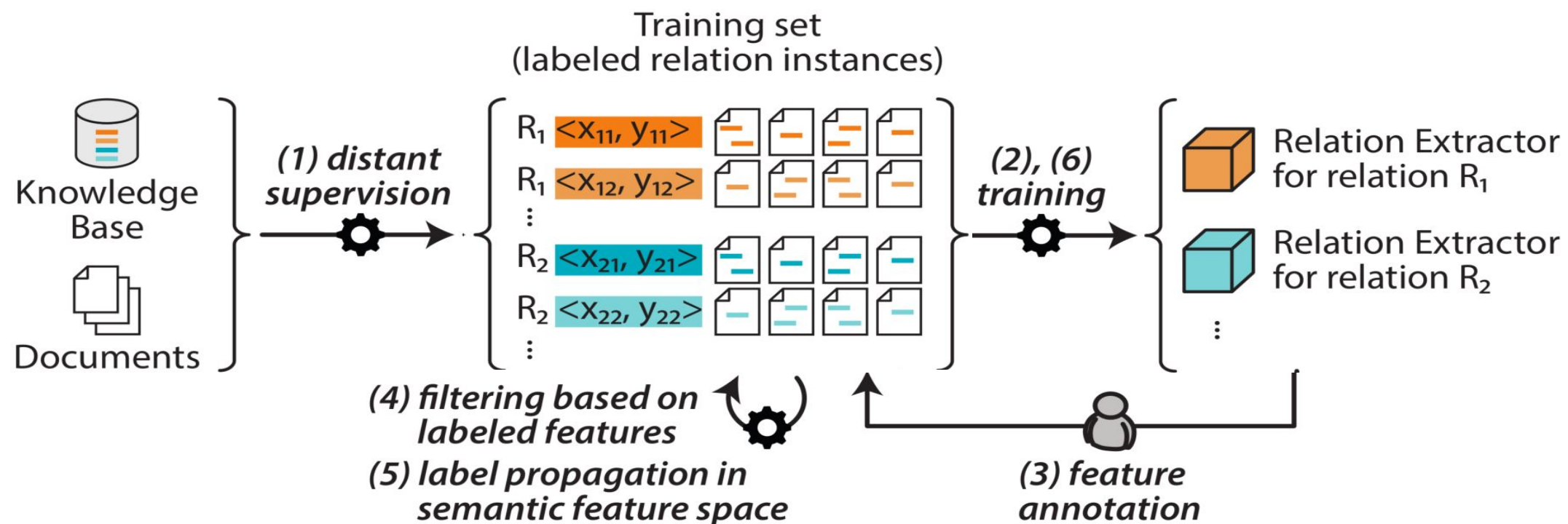
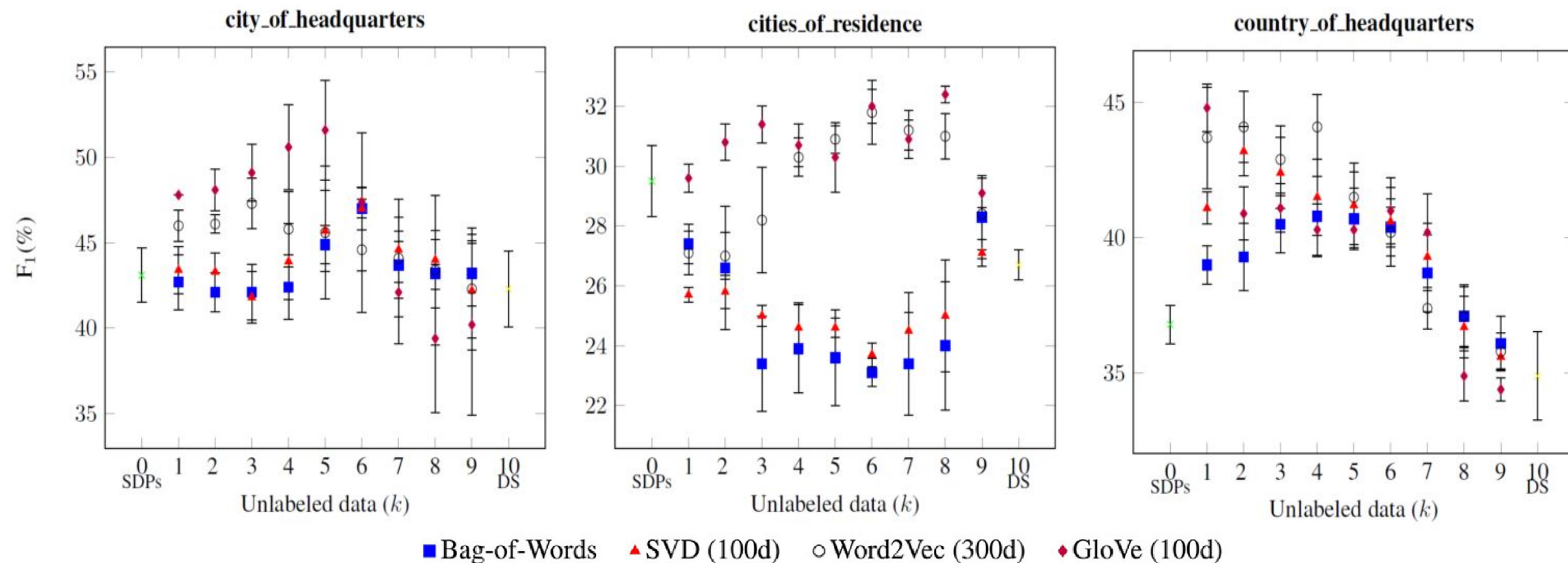
PER  $\xleftarrow{\text{appos}}$  executive  $\xrightarrow{\text{prep\_of}}$  ORG

$$\vec{C} = \text{CBOW}(\text{executive, of})$$

$$\text{Sim}(\vec{C}_t, \vec{C}_{DS}) = \frac{\vec{C}_t \cdot \vec{C}_{DS}}{|\vec{C}_t| \cdot |\vec{C}_{DS}|}$$

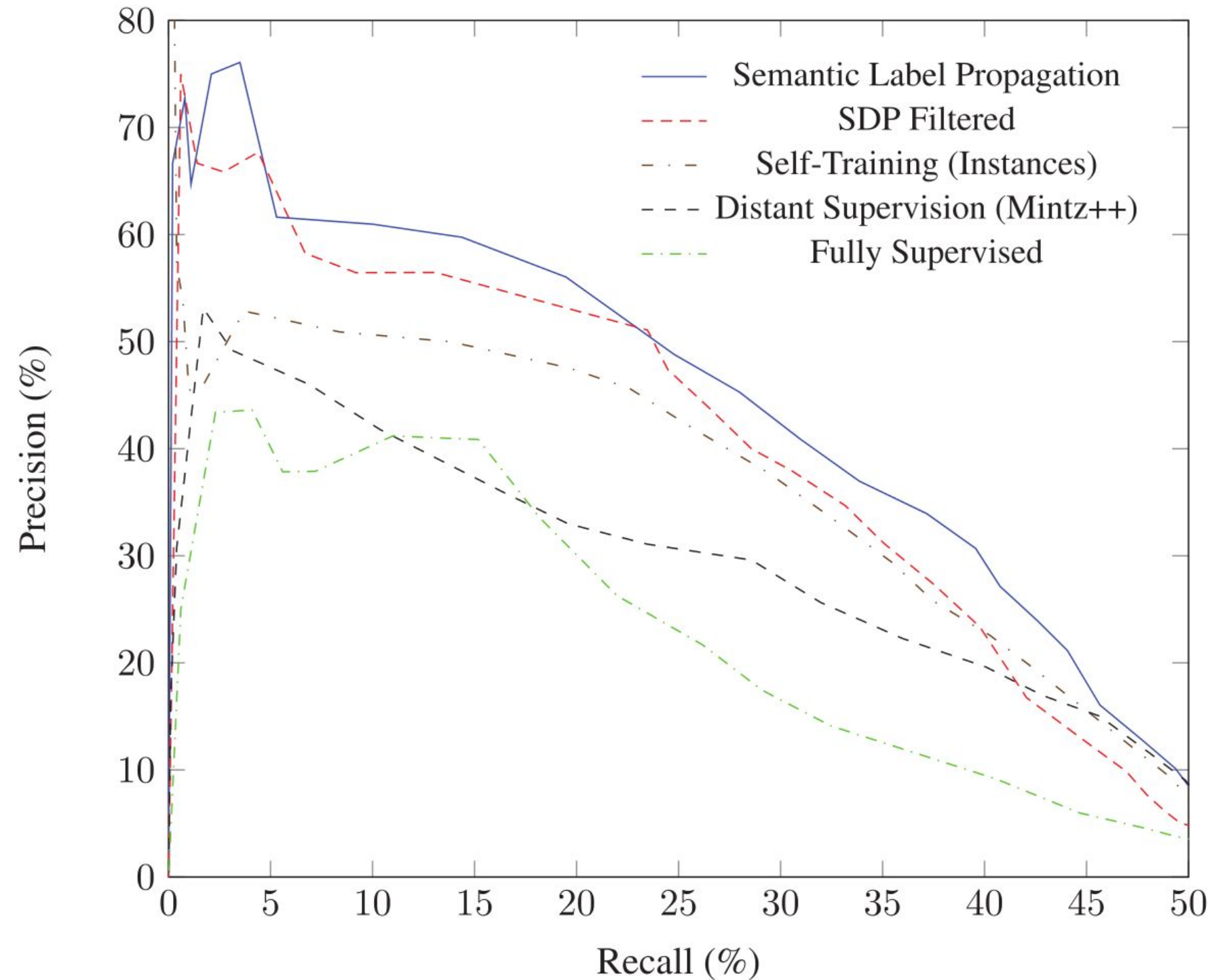


# Guiding Bootstrapped Relation Extractors

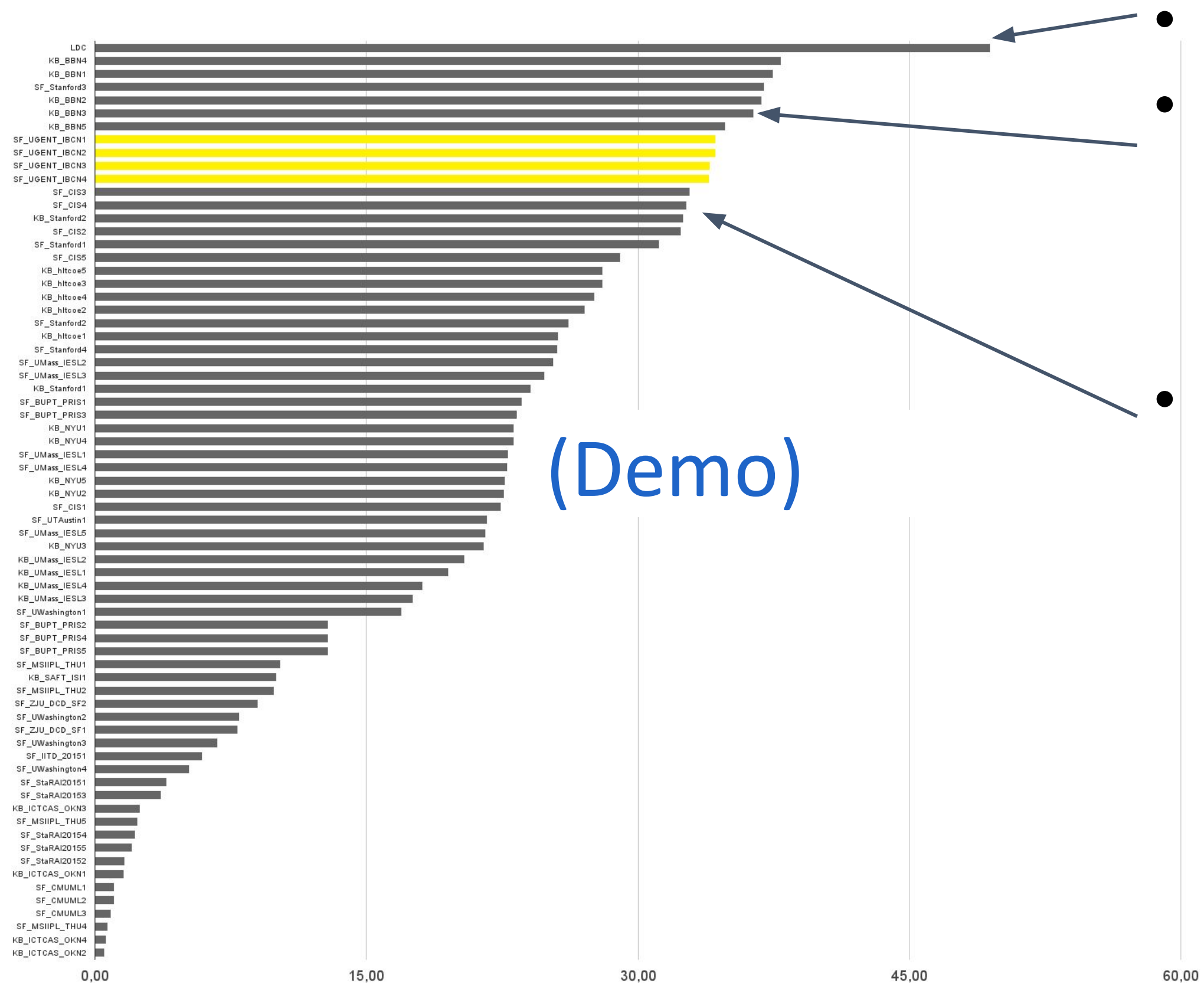


# Ghent University at TAC KBP

- **Recall** ↑↑
- **Minimal** supervision (5 min. per relation, 2u30 for TAC-KBP)



# Ghent University at TAC KBP

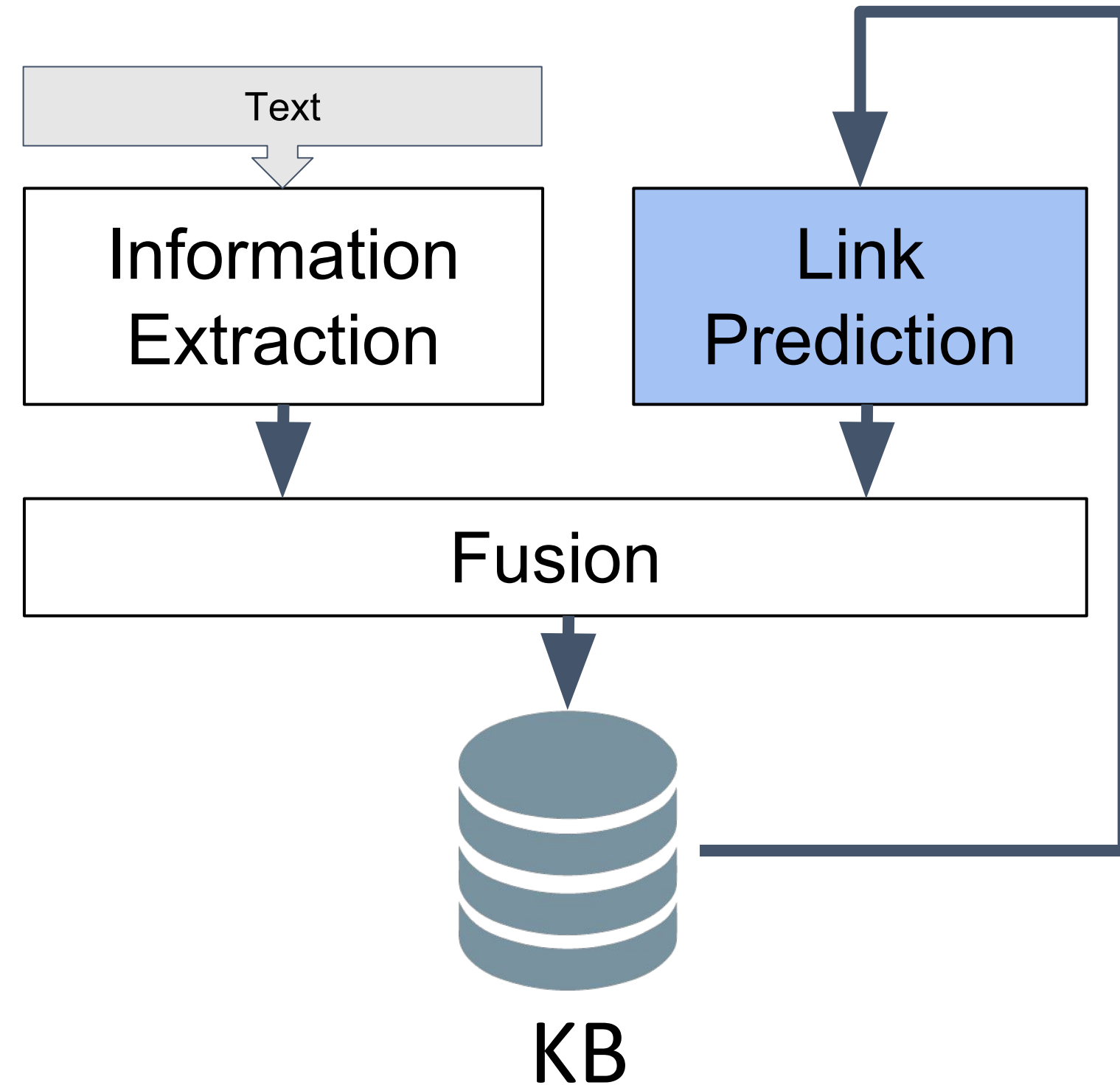


- Human
- DeepDive (Stanford)
  - ~40 hours total writing patterns
  - Supercomputer (786 Gig RAM)
- Ensemble of feature-based classifiers  
neural architectures, patterns,...

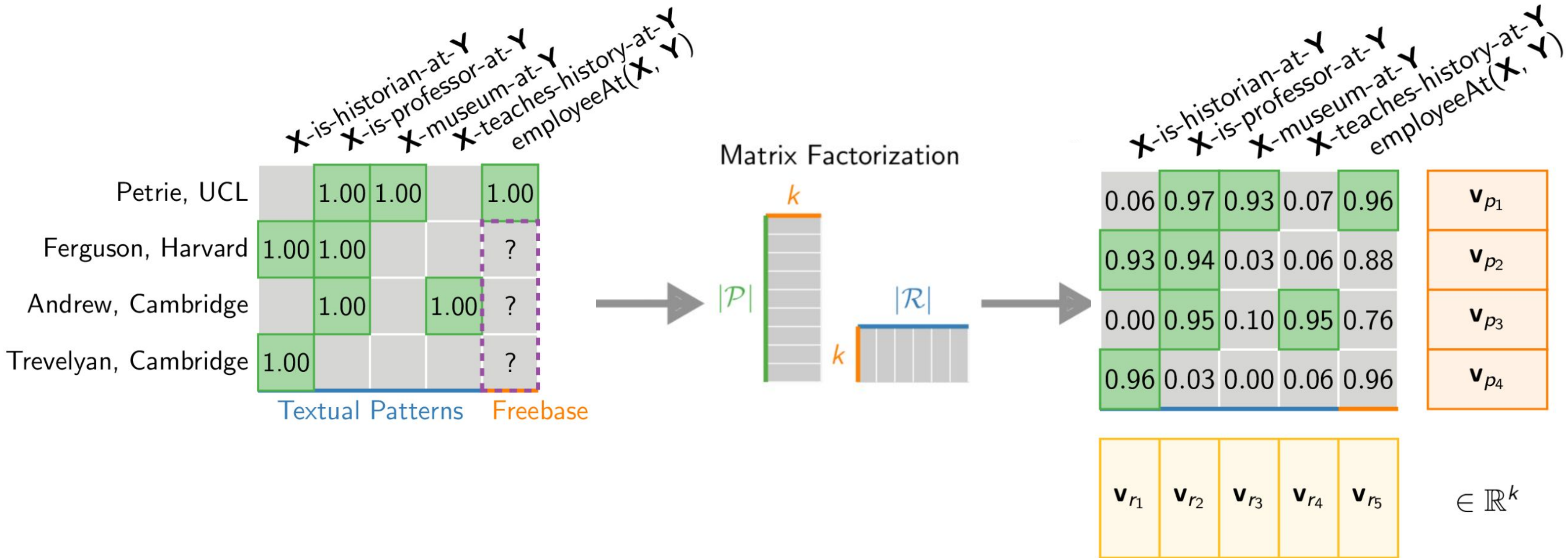


# Outline

1. Knowledge Bases
- 2. Knowledge Base Population**
  - a. Knowledge Extraction from Text - TAC KBP
  - b. Link Prediction**
3. Research Topic at Cambridge



# Relation Extraction by Matrix Factorization (Riedel, 2013)



$$p(\text{fact}) = p(x_{ij} = 1 \mid \text{Graph})$$

[neural] vector representations,

- + Similarity, approximate inference
- Fails for little alignment, hard to fix mistakes

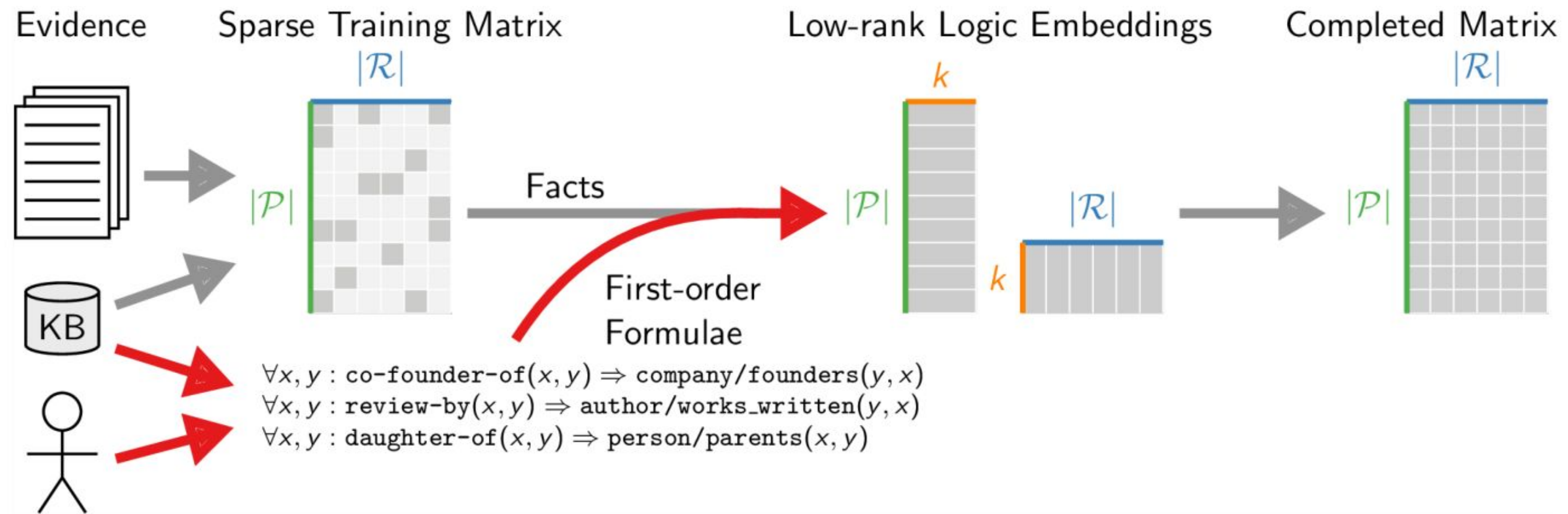
[symbolic] efficient (“lifted”) injection of prior knowledge    “prof\_at  $\Rightarrow$  works\_for”

- + Easy to modify
- Brittle, no generalization

combine **neural** and **symbolic** representations

to leverage advantages of both

# Injecting Logical Formulae (Rocktäschel, 2015)



- Inject general 1st order formulae
  - expressed in terms of probabilities of all training facts
  - e.g. model for  $r_p \Rightarrow r_q$  : by grounding over entities

$$p((r_p, e) \Rightarrow (r_q, e)) \approx 1 - p(r_p, e)(1 - p(r_q, e))$$

- Lessons learned:
  - + joint training of facts and rules works best
  - due to grounding, only practical for few rules

## Lifted implication rules (Demeester, 2016)

When is rule “`prof_at`  $\Rightarrow$  `works_for`” satisfied?

$$\forall e \in \mathcal{E} : p(\text{prof\_at}(e)) \leq p(\text{works\_for}(e))$$

$$\sigma(\mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e) \leq \sigma(\mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e)$$

$$\mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$$

“compatibility”

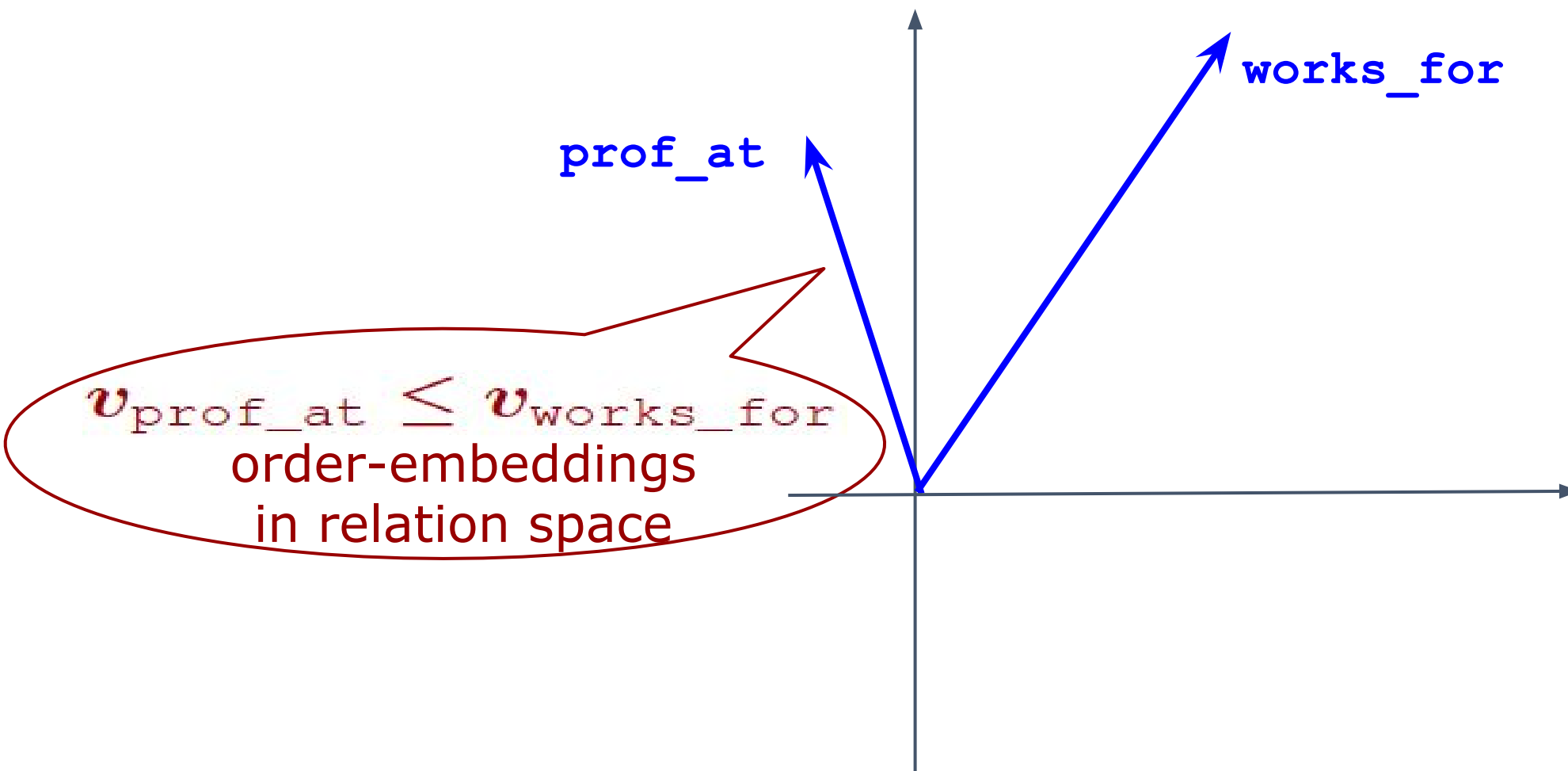
Sufficient (even stricter) condition:

$$\begin{cases} \mathbf{v}_{\text{prof\_at}} \leq \mathbf{v}_{\text{works\_for}} & \text{ordered relation embeddings} \\ \forall e \in \mathcal{E} : \mathbf{v}_e \in \mathbb{R}^{k,+} & \text{non-negative tuple embeddings} \end{cases}$$

# Lifted implication rules - illustration

rule `prof_at`  $\Rightarrow$  `works_for`

becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$



# Lifted implication rules - illustration

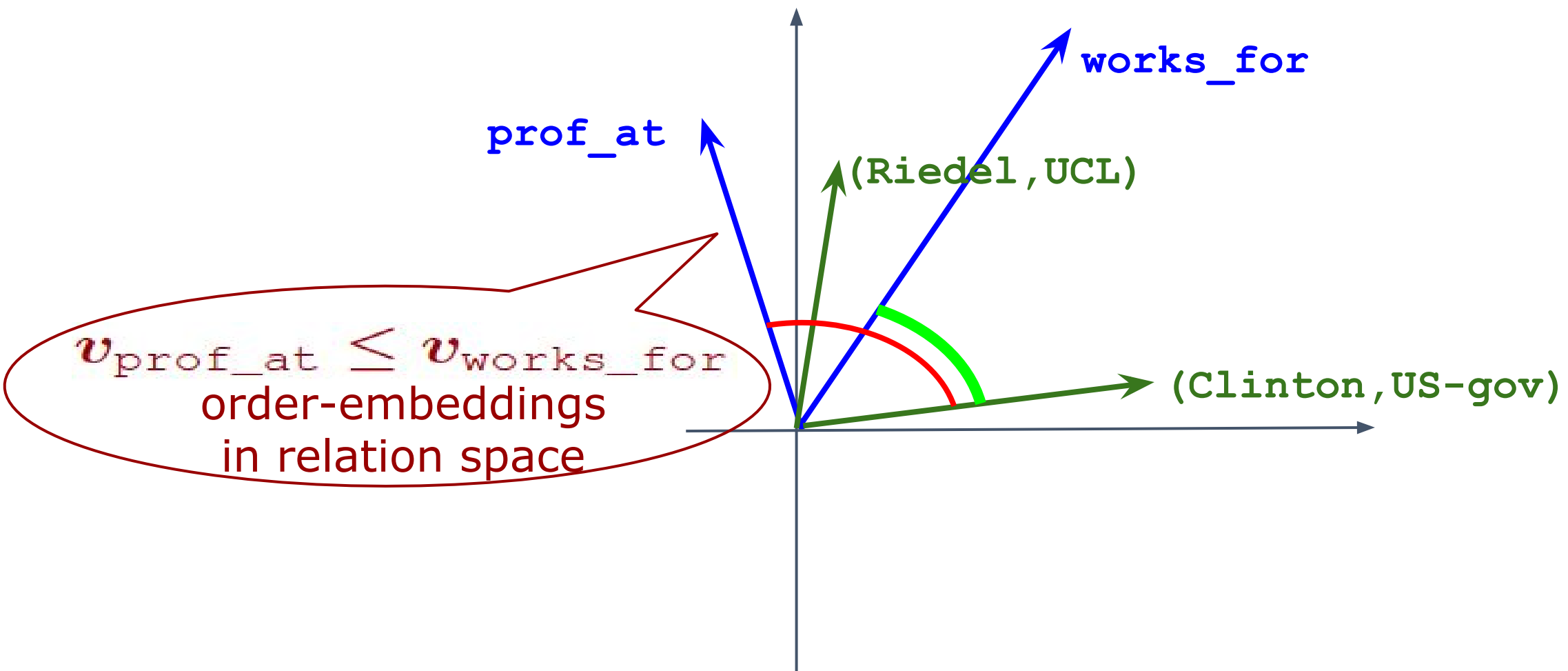
rule `prof_at`  $\Rightarrow$  `works_for`

becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$

Given: training facts

`works_for`(Clinton, US-Gov)

`prof_at`(Riedel, UCL)



# Lifted implication rules - illustration

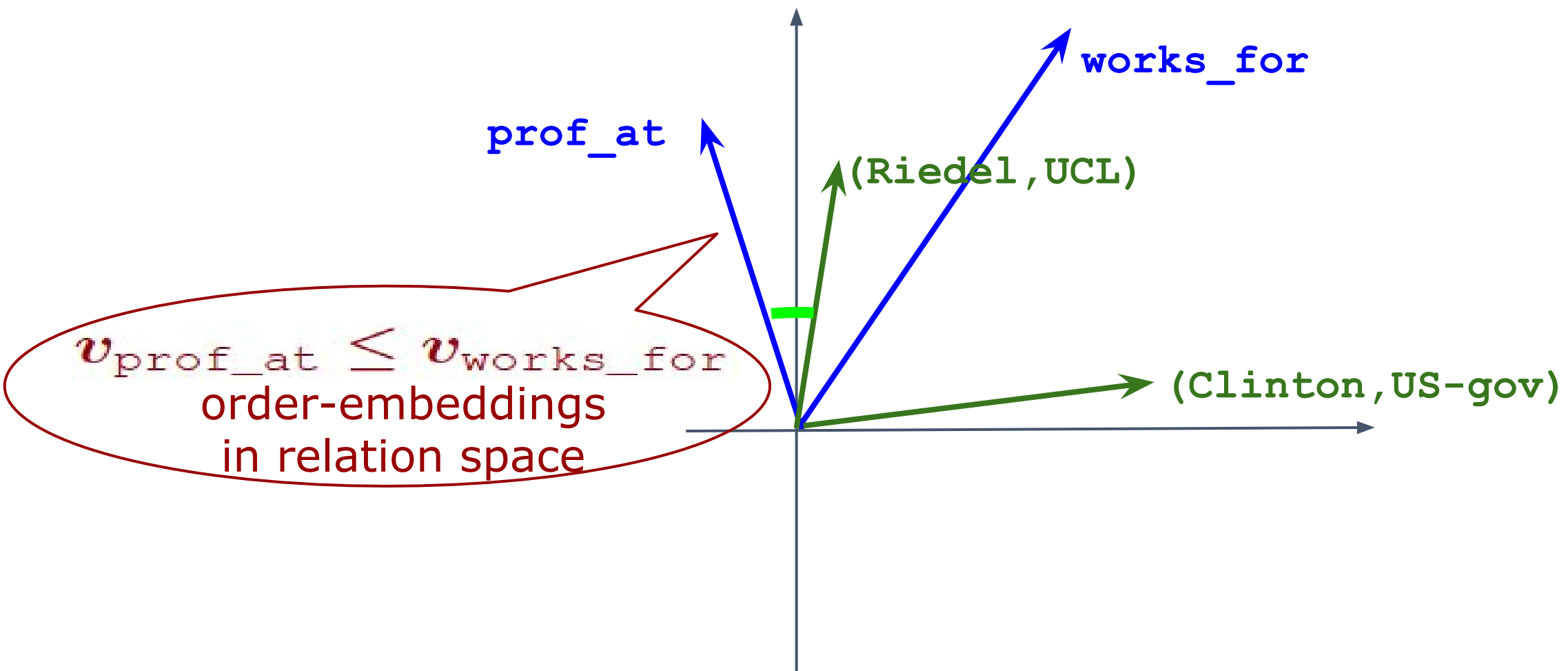
rule `prof_at`  $\Rightarrow$  `works_for`

becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$

Given: training facts

`works_for`(Clinton, US-Gov)

`prof_at`(Riedel, UCL)





# Lifted implication rules - illustration

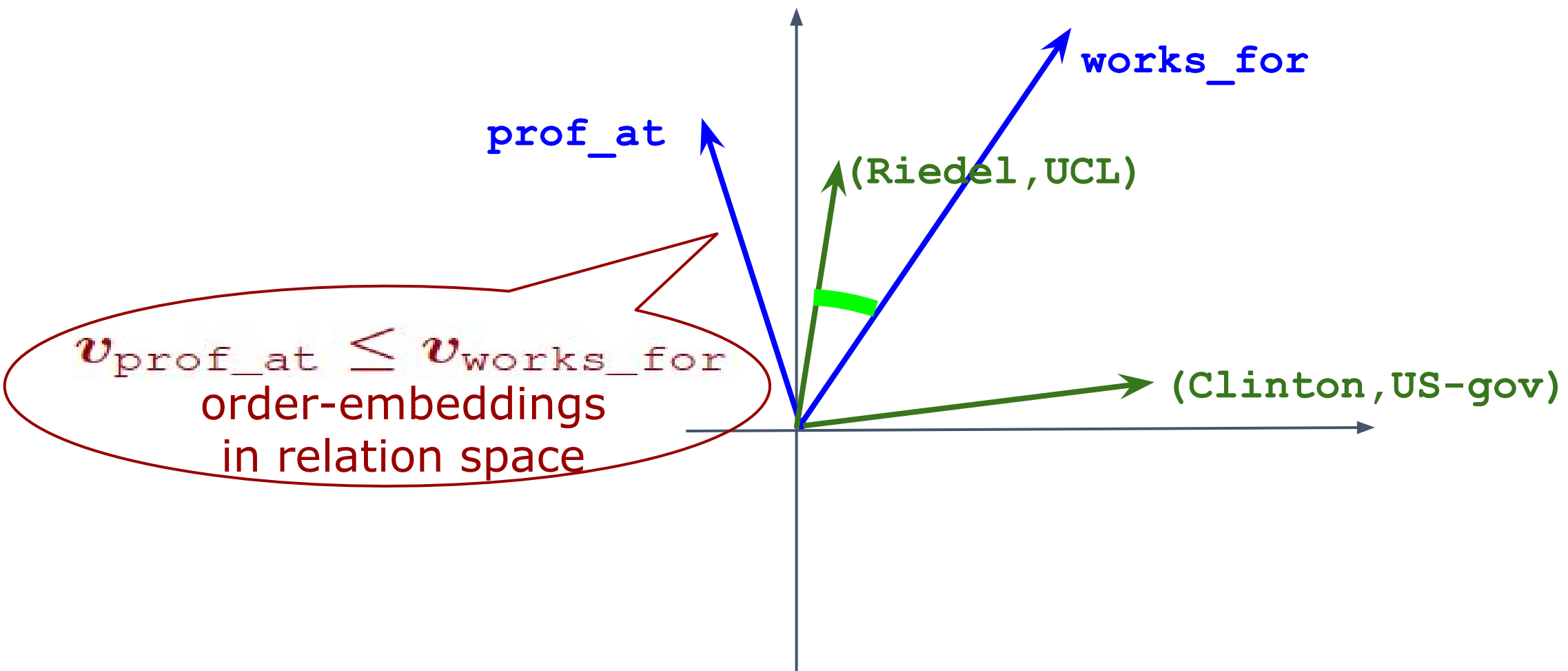
rule `prof_at`  $\Rightarrow$  `works_for`

becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$

Given: training facts

`works_for`(Clinton, US-Gov)

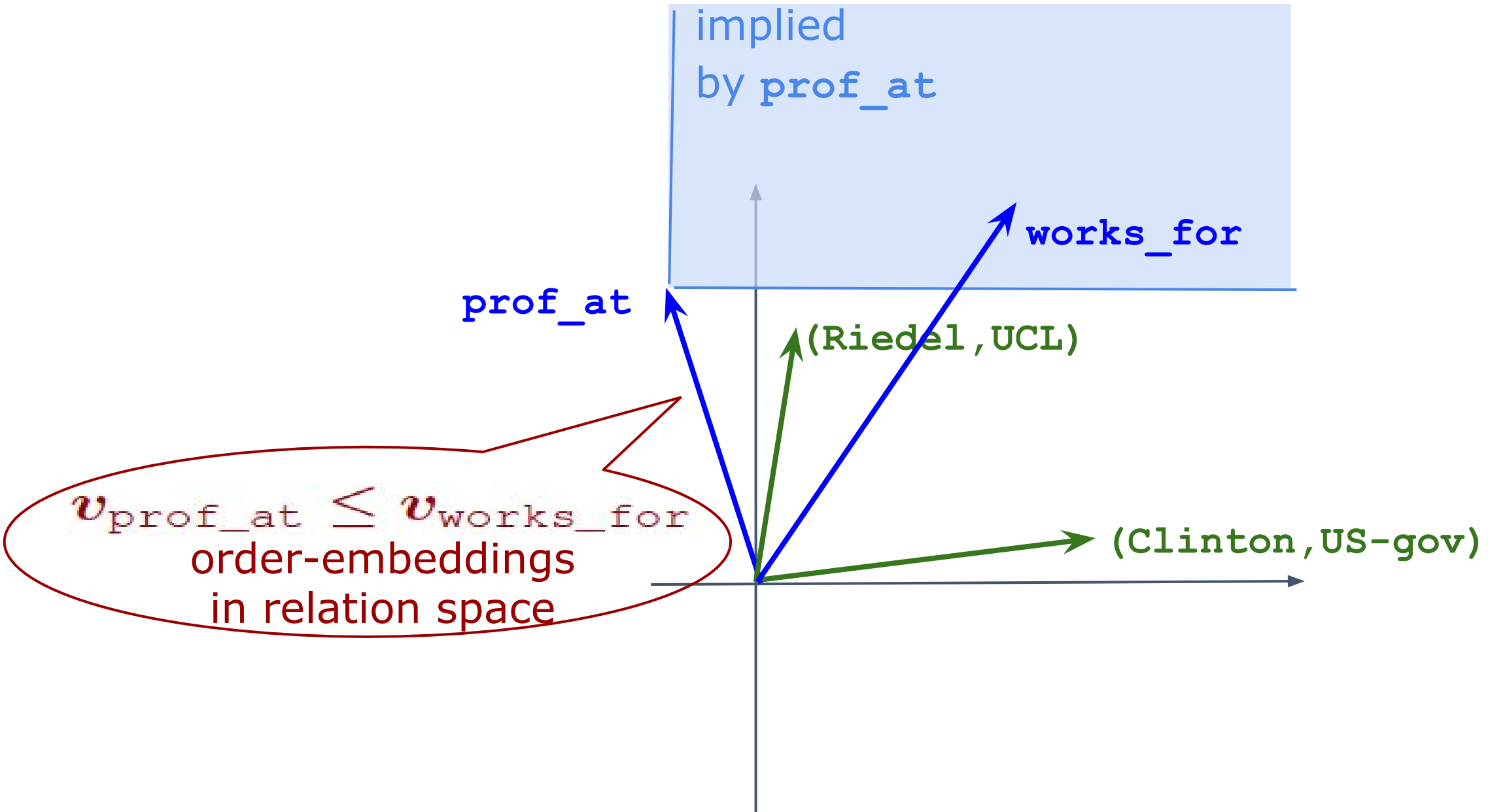
`prof_at`(Riedel, UCL)



# Lifted implication rules - illustration

rule `prof_at`  $\Rightarrow$  `works_for`

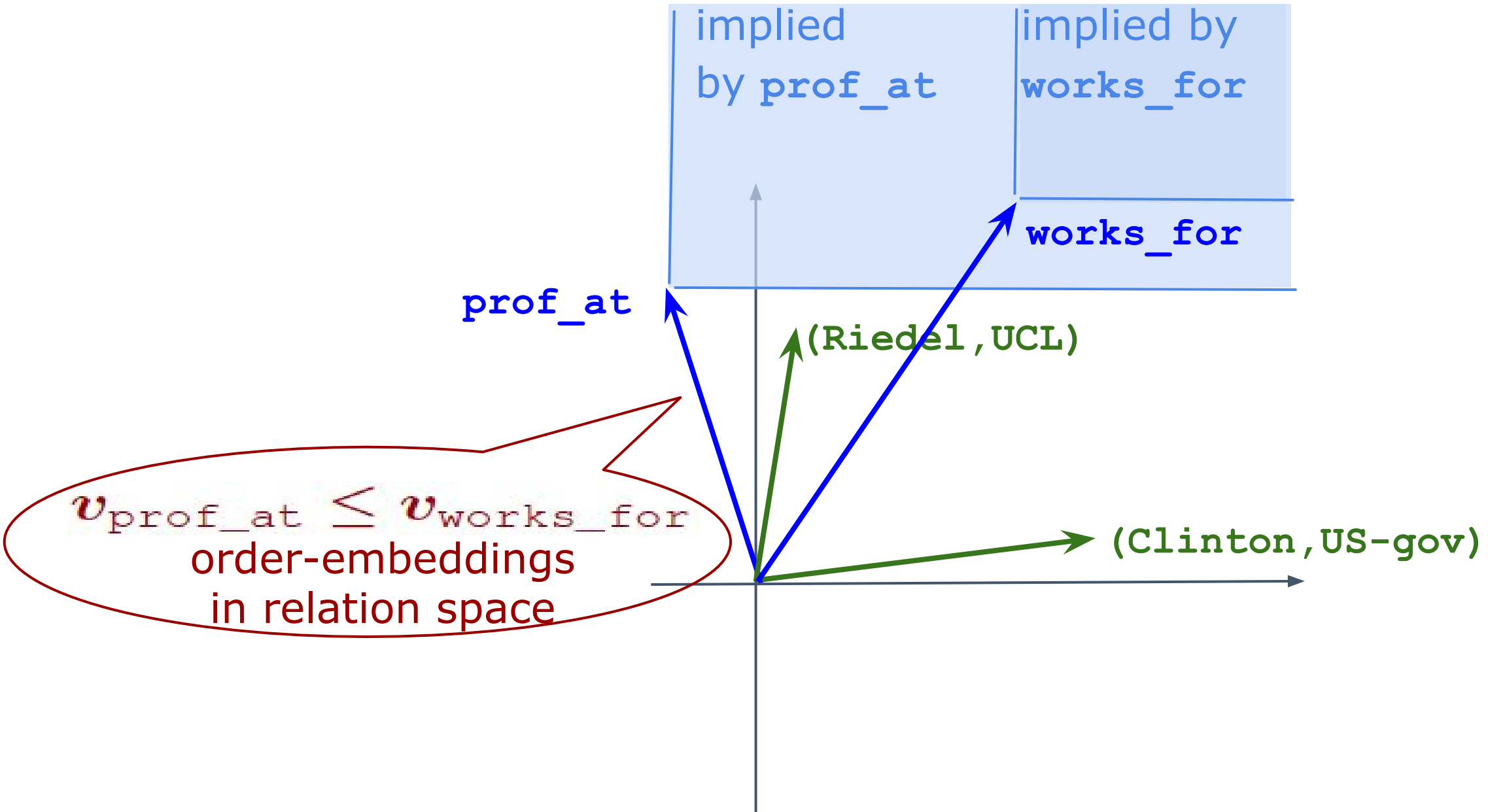
becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$



# Lifted implication rules - illustration

rule `prof_at`  $\Rightarrow$  `works_for`

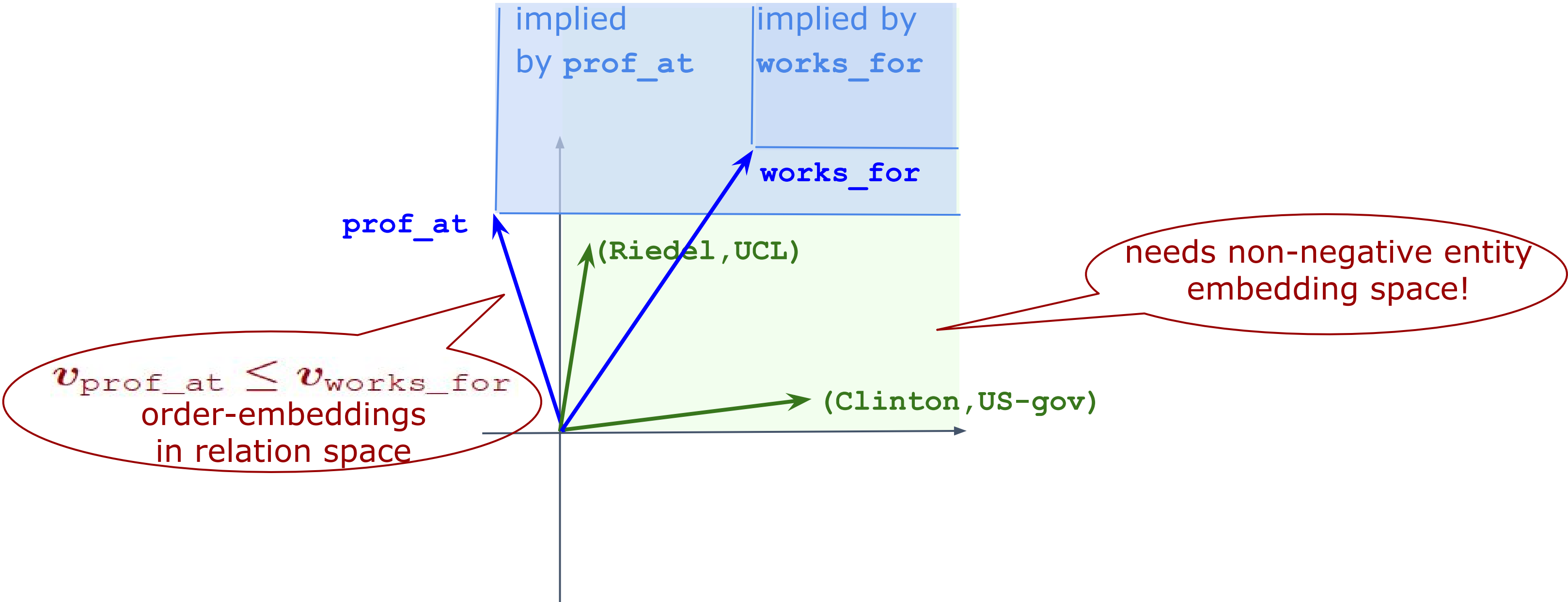
becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$



# Lifted implication rules - illustration

rule `prof_at`  $\Rightarrow$  `works_for`

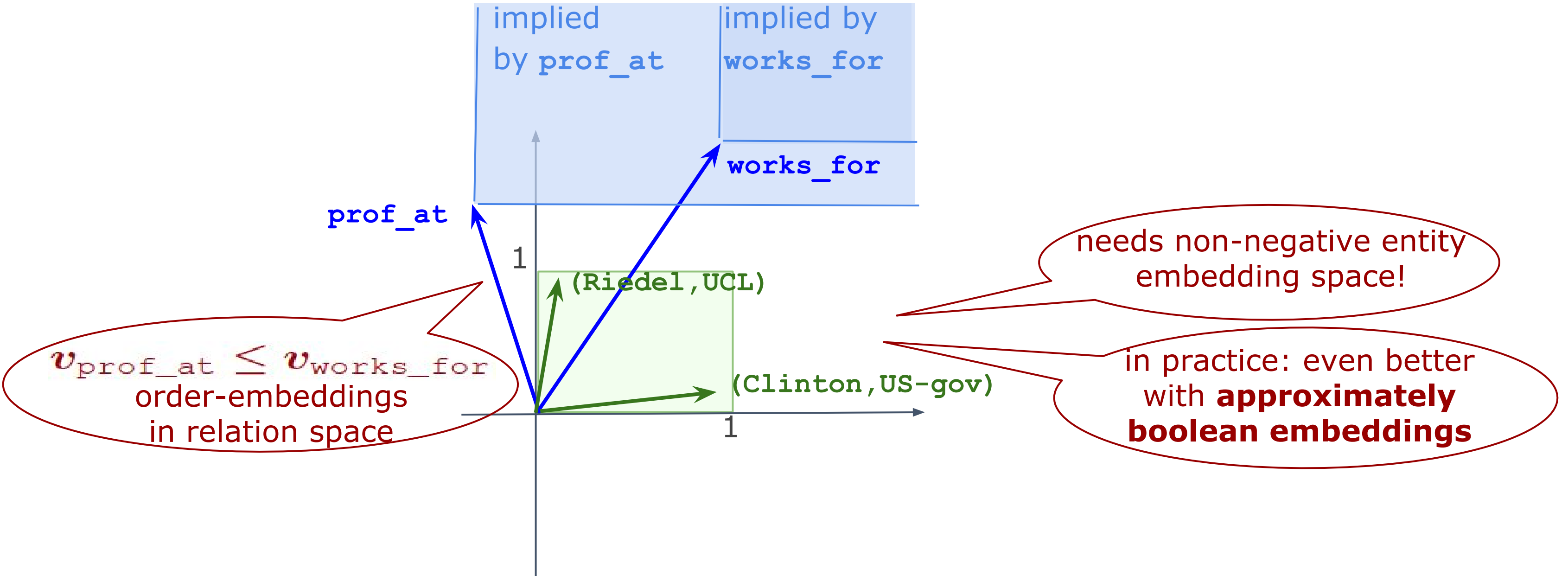
becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$



# Lifted implication rules - illustration

rule `prof_at`  $\Rightarrow$  `works_for`

becomes:  $\forall \mathbf{v}_e : \mathbf{v}_{\text{prof\_at}} \cdot \mathbf{v}_e \leq \mathbf{v}_{\text{works\_for}} \cdot \mathbf{v}_e$



# Lifted implication rules - in practice

## Non-negative entity embeddings?

Differentiable mapping of  $e \in \mathbb{R}^k$  to  $\tilde{e} \in \mathbb{R}^{k,+}$

Options:

- $\tilde{e} := \exp(\mathbf{e}) \in \mathbb{R}^{k,+}$
- $\tilde{e} := \text{ReLU}(\mathbf{e}) \in \mathbb{R}^{k,+}$
- $\tilde{e} := \sigma(\mathbf{e}) \in (0, 1)^k$

strongest restriction, but works best!  
"Approximately Boolean embeddings"

## Ordered relation embeddings?

1 additional "lifted" loss term per implication rule:

$$\text{minimize } \mathcal{L}_{\text{rule}} = \sum_i \max(0, [\mathbf{v}_{\text{prof\_at}} - \mathbf{v}_{\text{prof\_at}}]_i)$$

upper bound to  
"grounded" loss

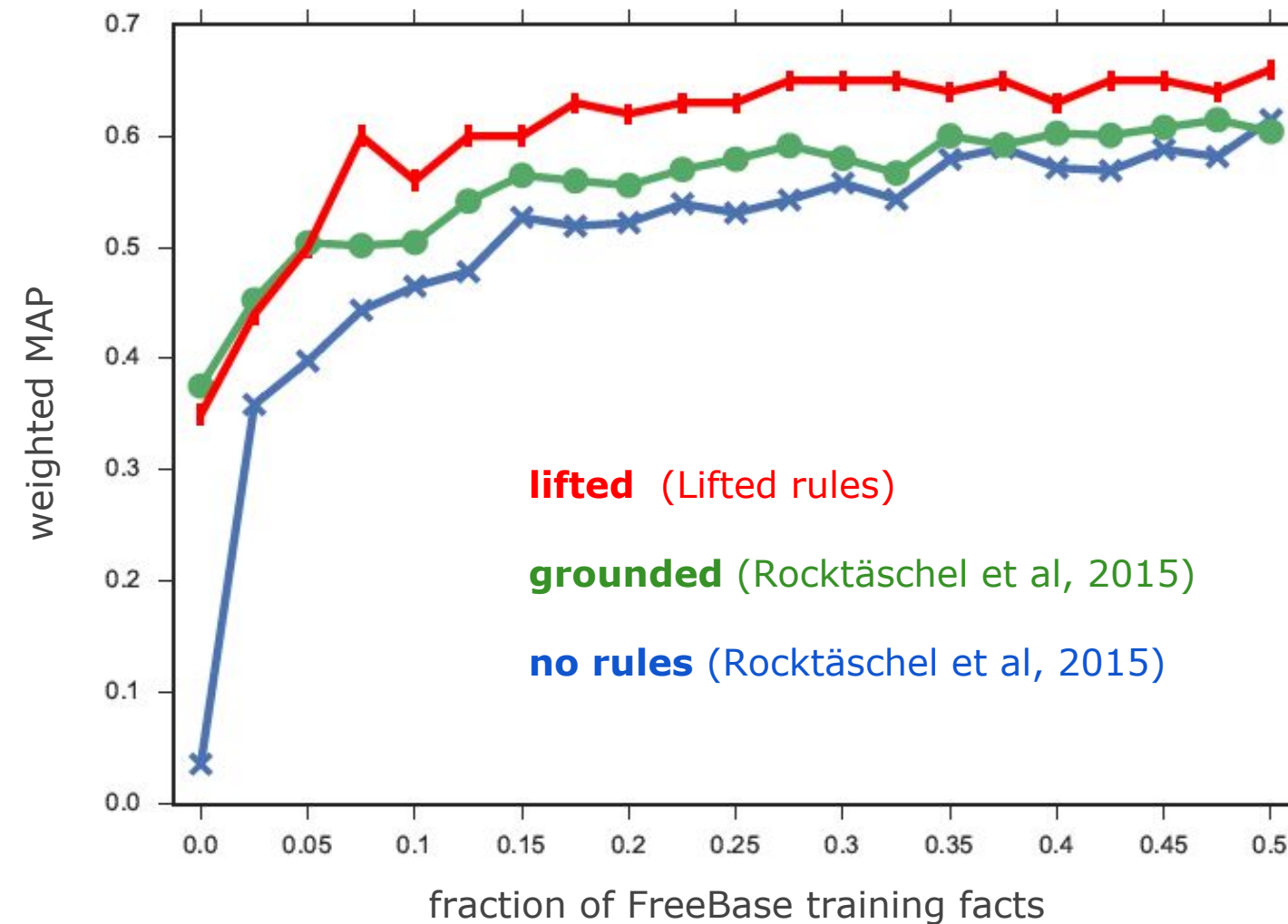
# Experiments: Grounded versus Lifted?

More efficient:

	0 rules	36 rules	427 rules
1 epoch (single CPU)	<b>6.33s</b>	<b>6.76</b>	<b>6.97s</b>

only  
10% overhead  
due to rules

Higher precision:



# Outline

1. Knowledge Bases
2. Knowledge Base Population
  - a. Knowledge Extraction from Text
  - b. Link Prediction
- 3. Research Topic at Cambridge**



# At Cambridge: Learning to Annotate

## Learn to explain hard-to-interpret text

- Genius.com
- 950.000 lyrics-annotation pairs
- English-to-English text generation
  - (Summarizing, Simplification, Paraphrasing,...)

Remember when I used to eat sardines for dinner  
 Peace to Ron G, Brucey B, Kid Capri  
 Funkmaster Flex, Lovebug Starski (wassup?)  
 I'm blowing up like you thought I would  
 Call the crib, same number, same hood (that's right)  
 It's all good (it's all good)  
 And if you don't know, now you know, nigga

[Hook: Total]  
 You know very well who you are  
 Don't let 'em hold you down, reach for the stars  
 You had a goal, but not that many  
 Cause you're the only one  
 I'll give you good and plenty

[Verse 2: The Notorious B.I.G.]  
 I made the change from a common thief  
 To up close and personal with Robin Leach  
 And I'm far from cheap, I smoke skunk with my peeps all day  
 Spread love, it's the Brooklyn way

### Genius Annotation 1 contributor

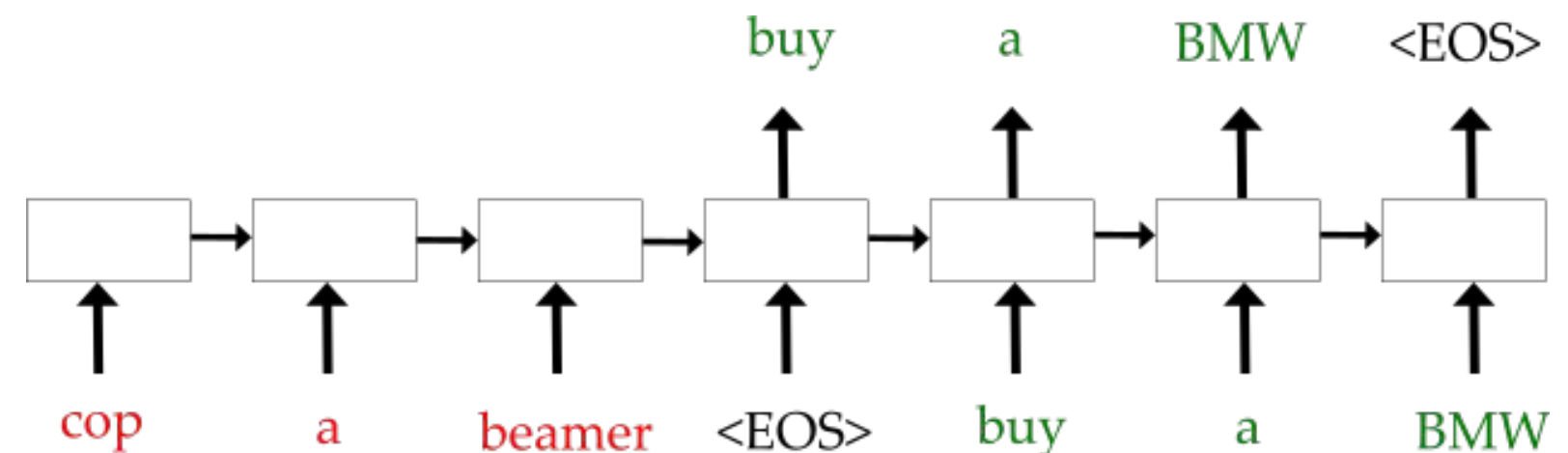
This iconic hook, sung by girl-group [Total](#), is a flip on the chorus of [Mtume's "Juicy Fruit"](#), which is sampled for the beat. The lyrics of the original are as follows:

“ You know very well what you are  
 You're my sugar thing, my chocolate star  
 I've had a few, but not that many  
 But you're the only love, that gives me good and plenty

Biggie's version flips the meaning, addressing his aspiration to fame rather than romantic love, but keeps the reference to [Good & Plenty candy](#), which B.I.G. presumably ate by the handful.



- Statistical and Neural Machine Translation?
- Translation vs. Retrieval based?
- Evaluation ?
- Paraphrasing vs. External Knowledge ?



# Learning to Annotate

---

Lyrics: roly on my arm  
True: he's always rocking a rolex  
Retr.: (5) he wears a rolex, rolex manufactures expensive wristwatches.  
LSTM: (4) he wears his rolex

---

Lyrics: i blast ya cabbage  
True: think "head of cabbage". he's gonna shoot you in the head.  
Retr.: (3) brains, post-headshot  
LSTM: (4) he ' s gonna shoot you .

---

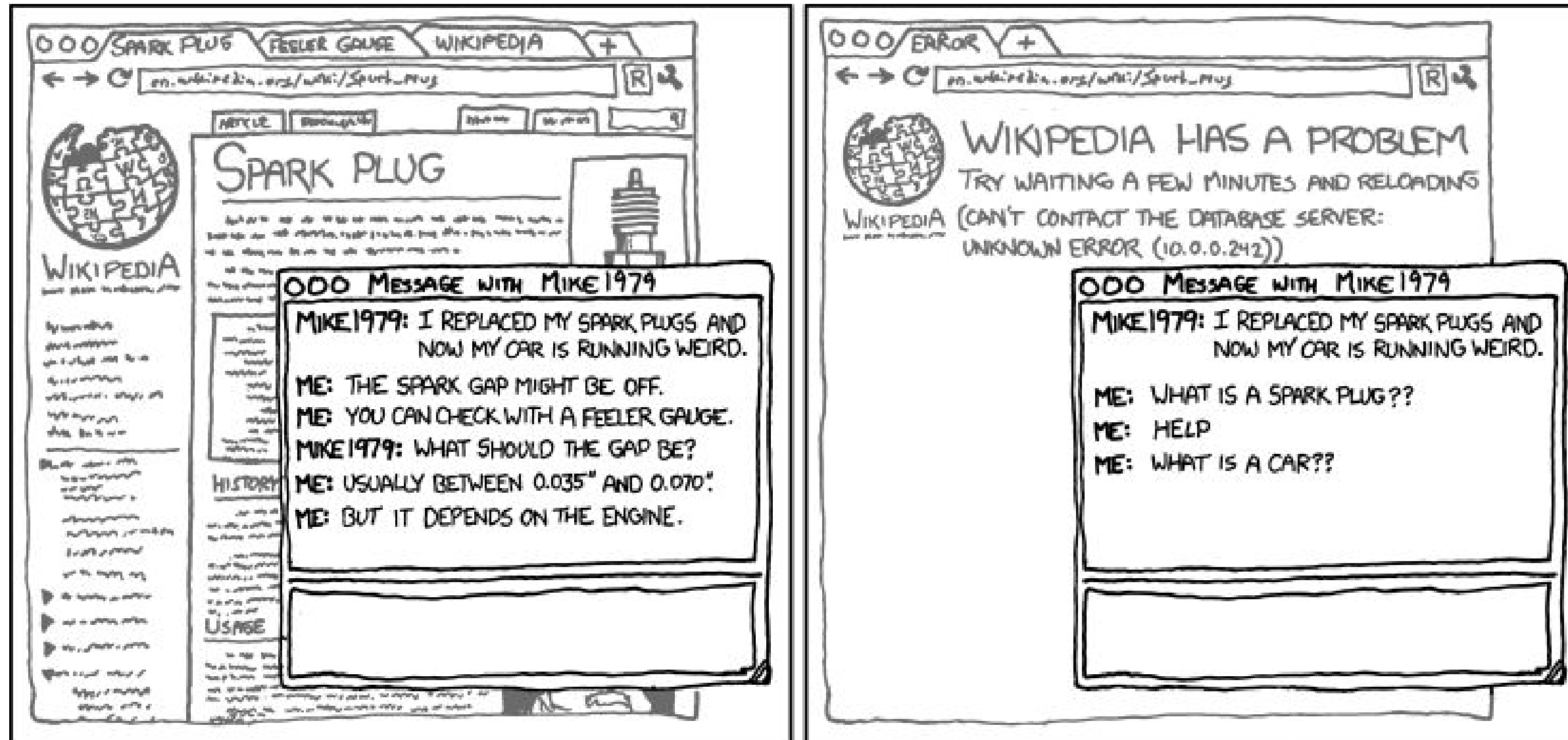
Lyrics: i used to have to pack a mack in back of the ac  
True: skits called "packin a mac in the back of the ac" appeared on pun's first two albums  
Retr.: (5) he's "packin a mac [pakinamac] in the back of the ac [meaning acura]  
vinnie is saying that he'll put you in the back of an accura right next to big pun's mac.  
LSTM: (4) he used to have a gun in the back of the car

---

Lyrics: behold the flyest; bentley drivers louis vuitton buyers, jet fuel abusers  
driving bentleys,  
True: buying louis vuitton swag, and burning fuel in private jets are ways to show the cash that biggie and jay-z have to spend.  
(they are the flyest)  
Retr.: (2) for rich people/ballers only, prestigious (expensive as hell) name brand.  
LSTM: (4) he ' s got a lot of money and expensive cars , and he ' s got a lot of expensive brands .

---

# Thank you ! Questions, Comments?

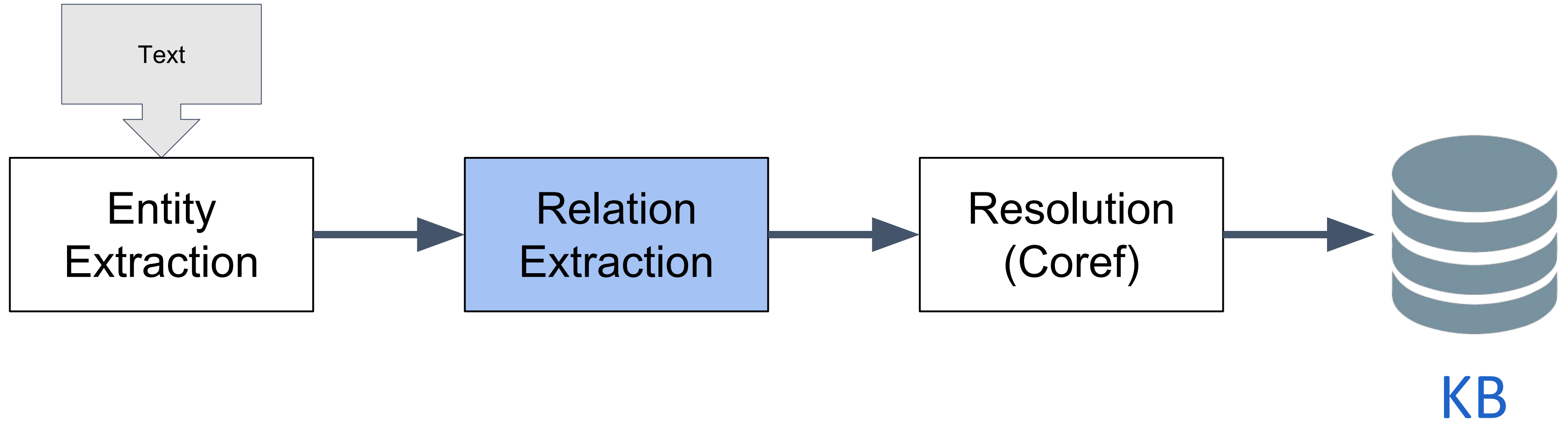


WHEN WIKIPEDIA HAS A SERVER OUTAGE, MY APPARENT IQ DROPS BY ABOUT 30 POINTS.

Xkcd extended mind

Lucas Sterckx  
lusterck.github.io  
@lusterck

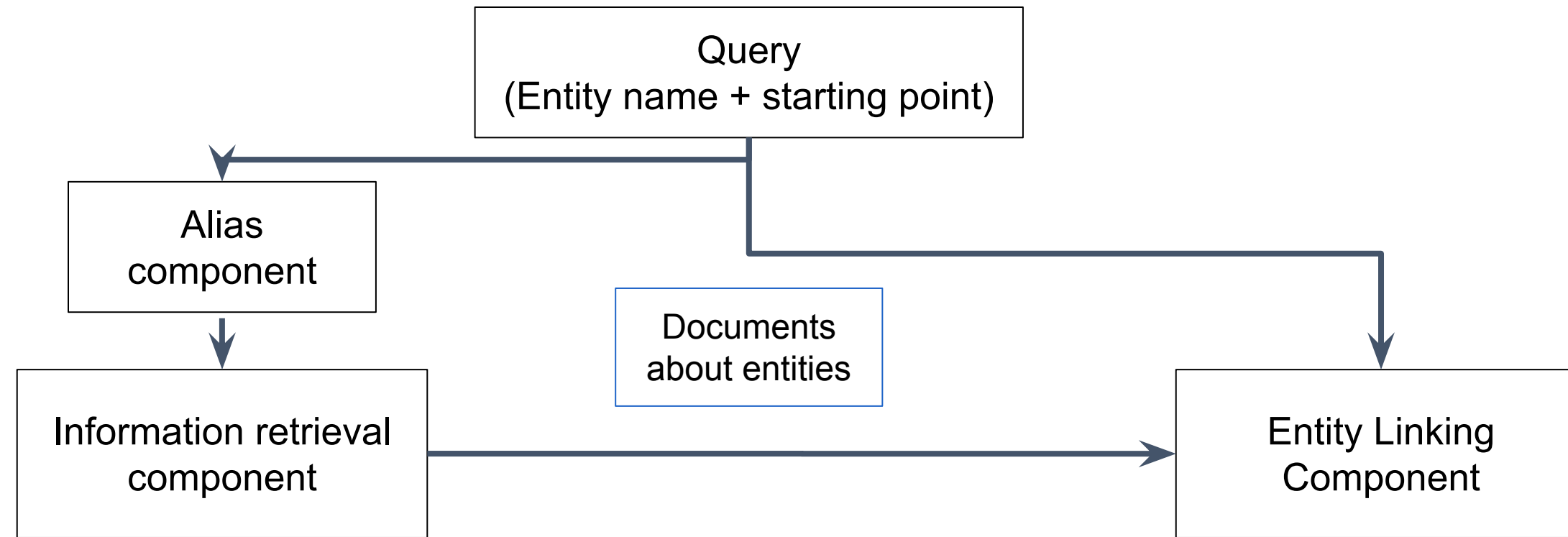
# Information Extraction



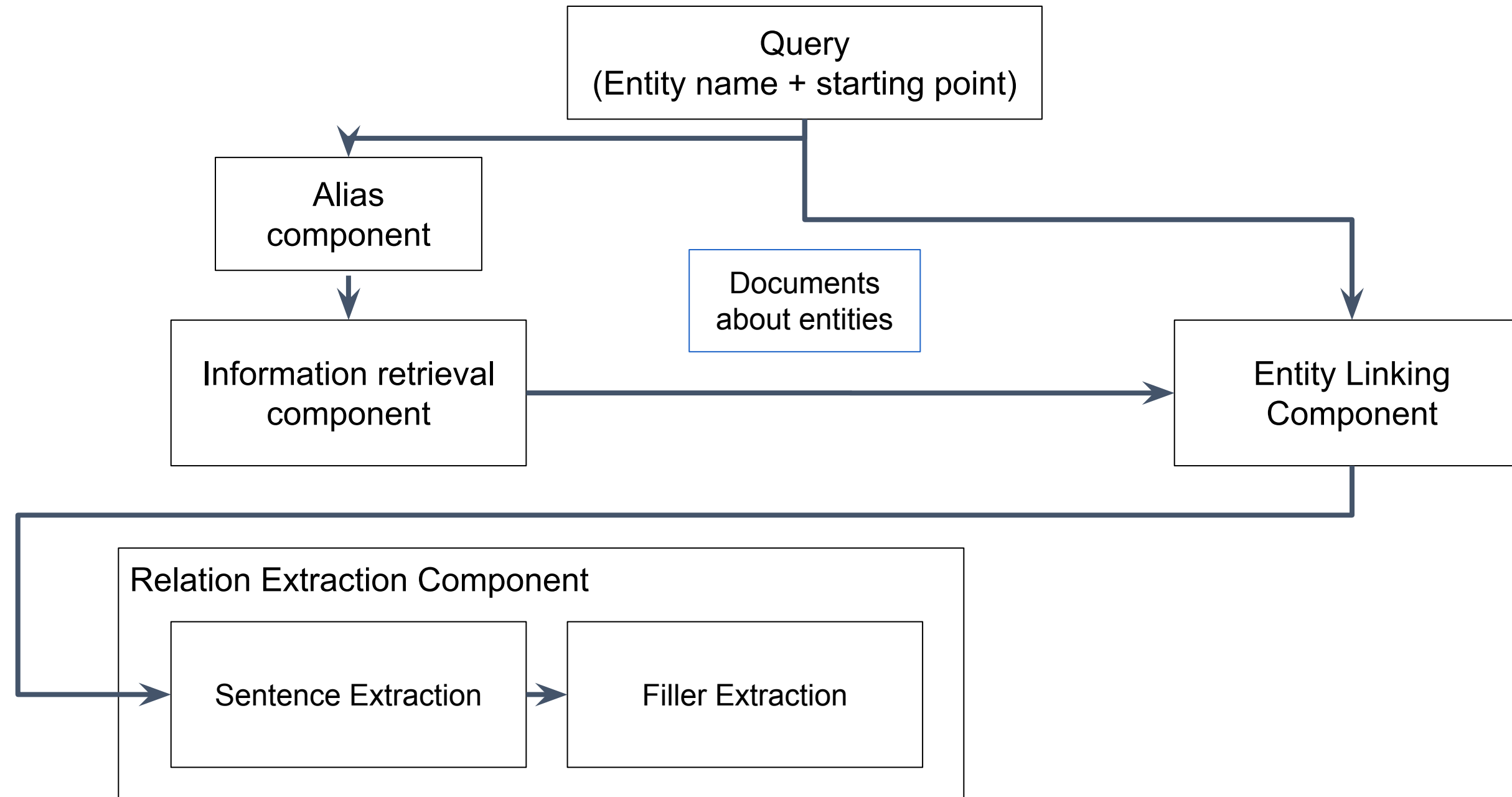
# Ghent University at TAC KBP

Query  
(Entity name + starting point)

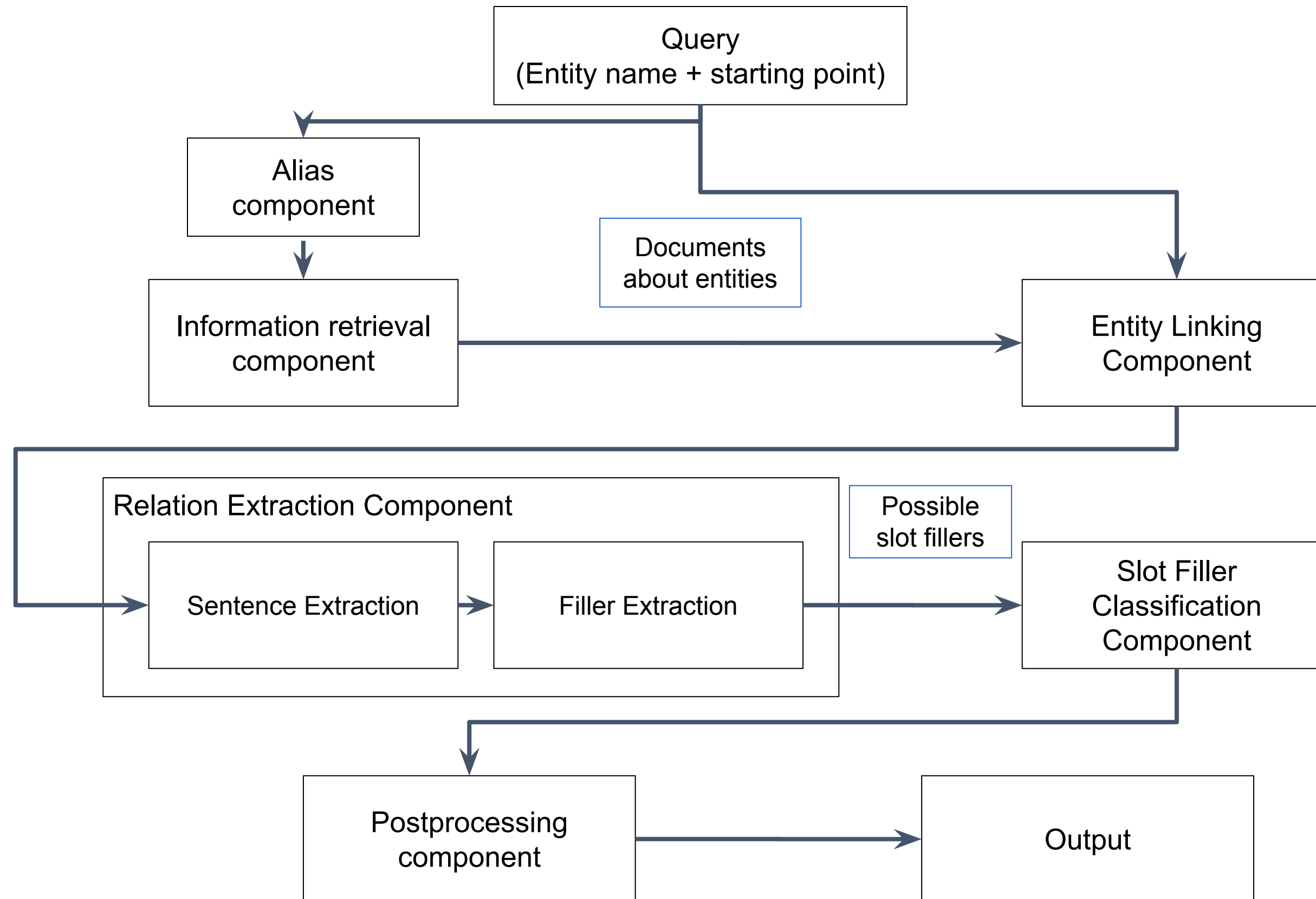
# Ghent University at TAC KBP



# Ghent University at TAC KBP



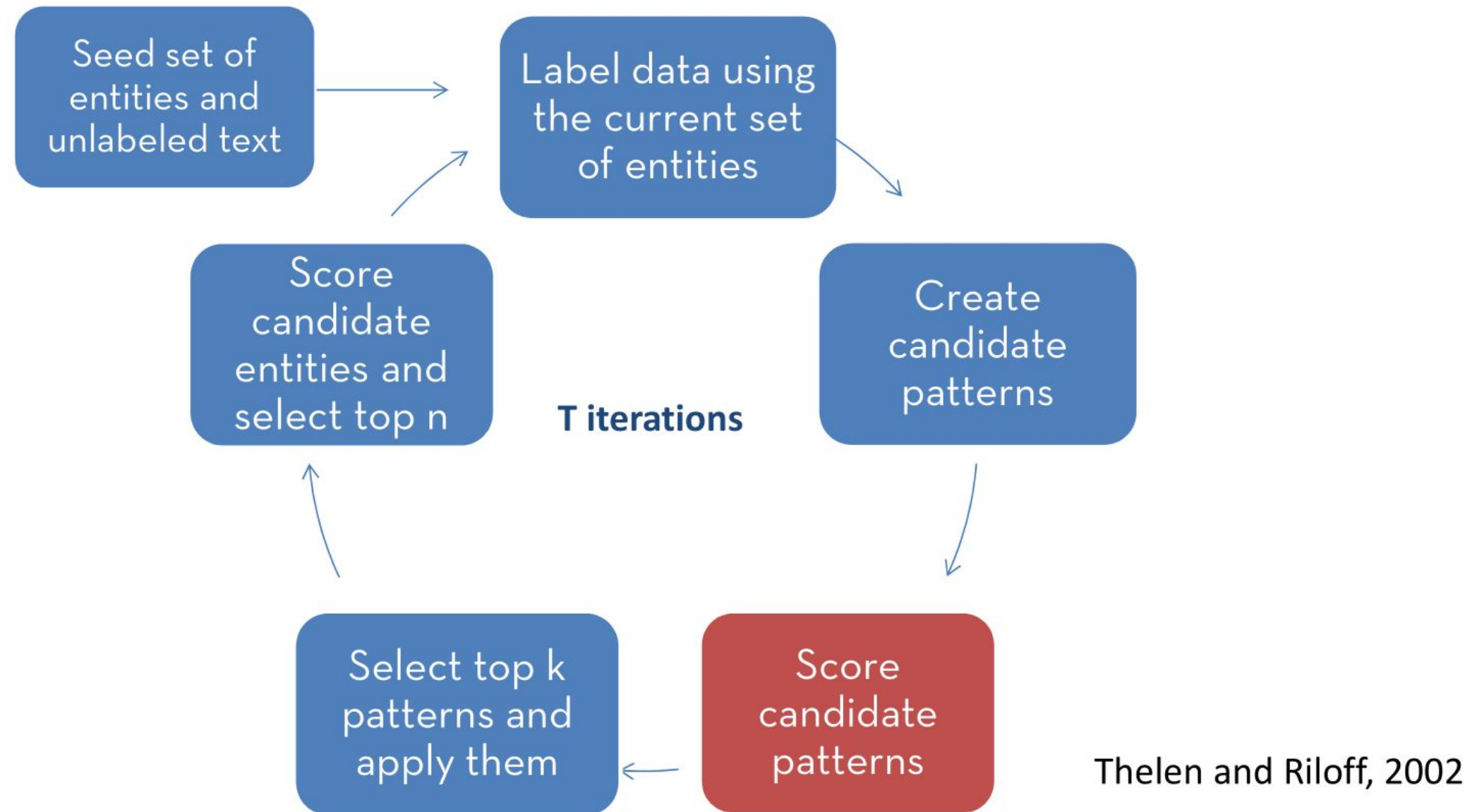
# Ghent University at TAC KBP





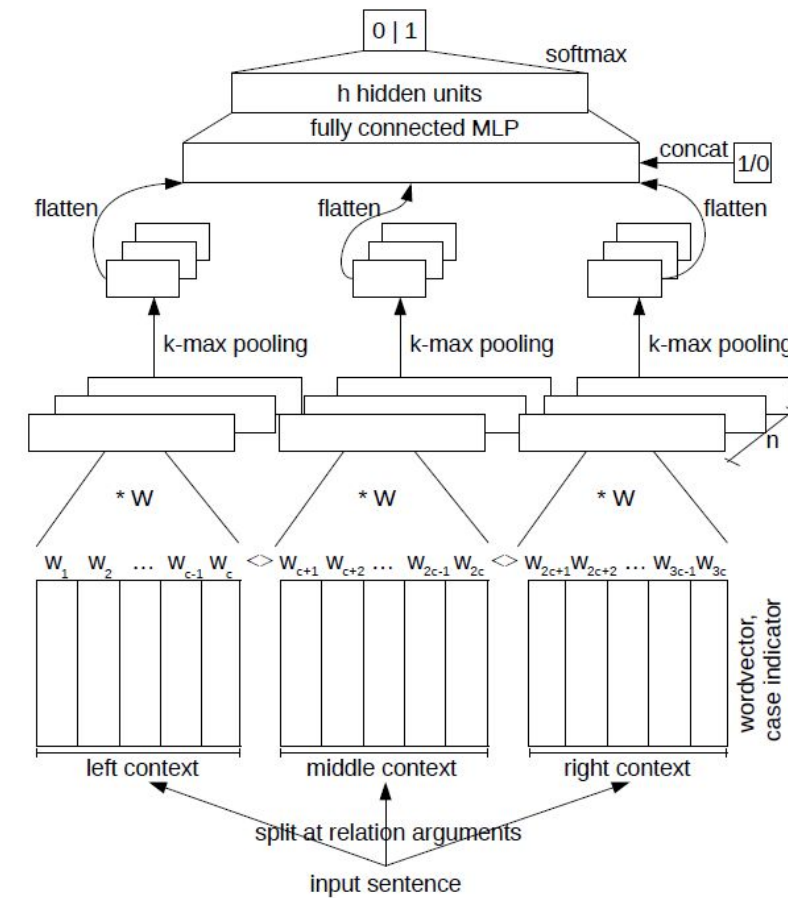
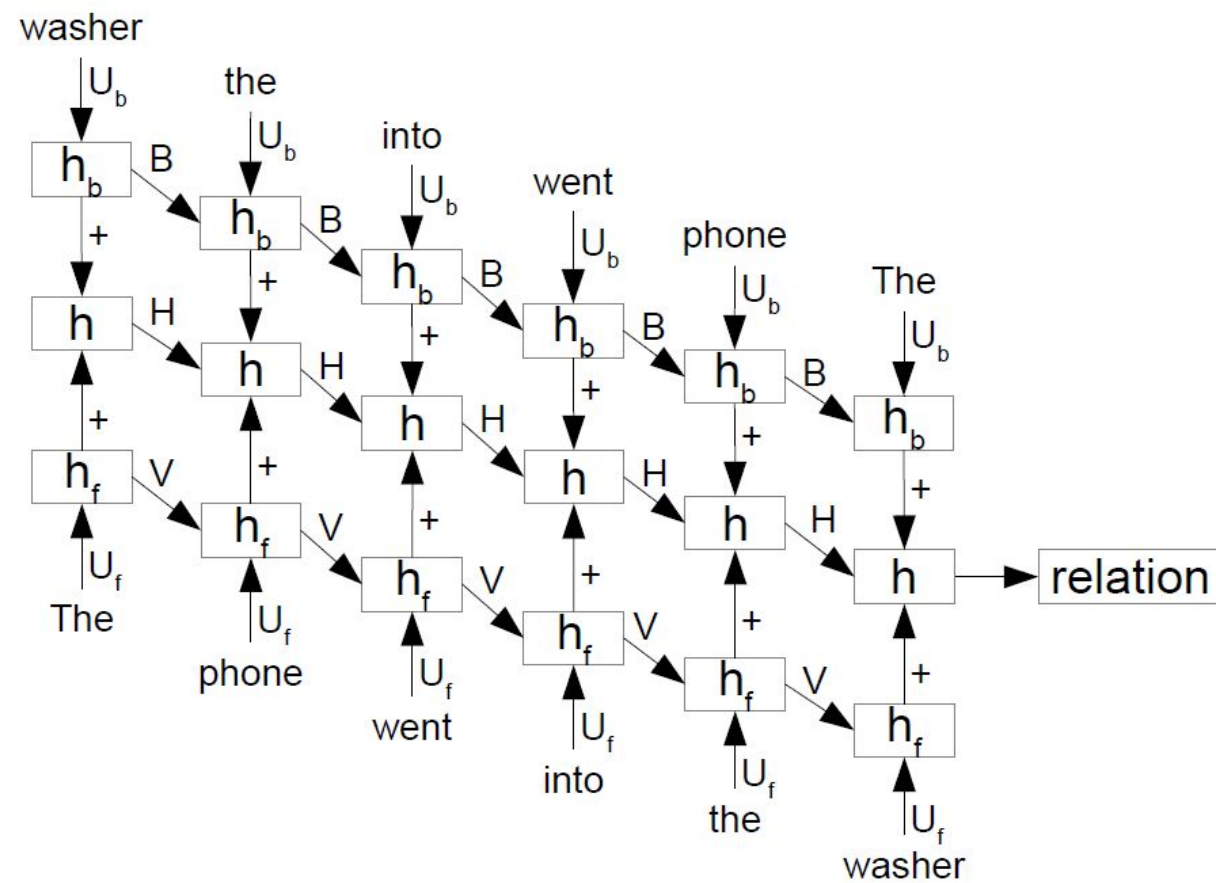
# Semi-Supervised Relation Extraction

→ **Bootstrapping** (Hearst, DIPRE, Snowball, BRED)



# Supervised Relation Extraction

- **LSTM or CNN-based** sentence classifiers



# Link Prediction in Knowledge Graphs

$$\Pr(x_{ijk} = 1 \mid \text{Graph})$$

	$x$ -is-historian-at- $Y$	$x$ -is-professor-at- $Y$	$x$ -museum-at- $Y$	$x$ -teaches-history-at- $Y$	$\text{employeeAt}(x, Y)$
Petrie, UCL	1.00	1.00		1.00	$\mathbf{v}_{p1}$
Ferguson, Harvard	1.00	1.00			$\mathbf{v}_{p2}$
Andrew, Cambridge	1.00		1.00		$\mathbf{v}_{p3}$
Trevelyan, Cambridge	1.00				$\mathbf{v}_{p4}$
	$\mathbf{v}_{r1}$	$\mathbf{v}_{r2}$	$\mathbf{v}_{r3}$	$\mathbf{v}_{r4}$	$\mathbf{v}_{r5}$

$\in \mathbb{R}^k$

# Knowledge Bases enable AI

- **“Knowledge is Power”**
  - “If a program is to perform a complex task well, it must know a great deal about the world in which it operates.”
- **Graphs** can be processed **efficiently** and offer a convenient abstraction
- **Enabling technology** for:
  - Machine reasoning
  - Disambiguation in written and spoken data
  - Semantic search in terms of entities & relations (not keywords & pages )

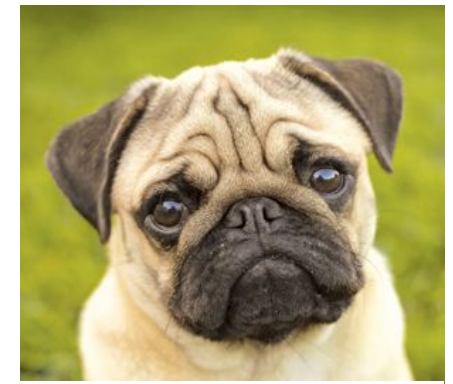
# Knowledge representations



Whisky



Tarzan



Snoopy

	<code>is_cat</code>	<code>is_dog</code>	<code>is_animal</code>
Whisky	1	0	1
Tarzan	1	0	1
Snoopy	0	1	1

# Knowledge representations

neural

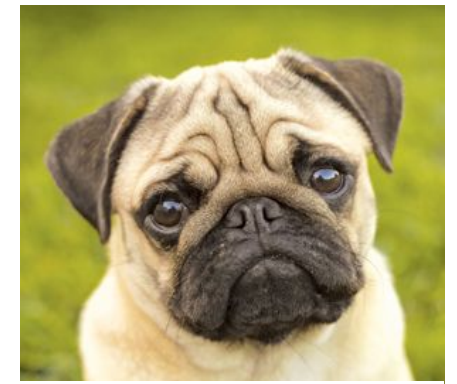
symbolic



Whisky



Tarzan



Snoopy

low-dimensional representations

- can capture similarity / hierarchy
- can be trained from raw facts
- Difficult to incorporate prior knowledge!

“all cats are animals”

**grounded** in  
all entities!

$$\mathbf{v}_{\text{Whisky}}, \mathbf{v}_{\text{is\_cat}}, \mathbf{v}_{\text{is\_animal}} \in \mathbb{R}^k$$

$$\mathbf{v}_{\text{Whisky}} \approx \mathbf{v}_{\text{Tarzan}}$$

$$p(\text{fact}) := \sigma(\mathbf{v}_{\text{predicate}} \cdot \mathbf{v}_{\text{entity}})$$

$$\sigma(\mathbf{v}_{\text{is\_cat}} \cdot \mathbf{v}_{\text{Tarzan}}) \approx 1$$

$$\sigma(\mathbf{v}_{\text{is\_cat}} \cdot \mathbf{v}_{\text{Whisky}}) \approx 1$$

$$\sigma(\mathbf{v}_{\text{is\_animal}} \cdot \mathbf{v}_{\text{Tarzan}}) \approx 1$$

$$\sigma(\mathbf{v}_{\text{is\_animal}} \cdot \mathbf{v}_{\text{Whisky}}) \approx 1$$

# Knowledge representations

neural

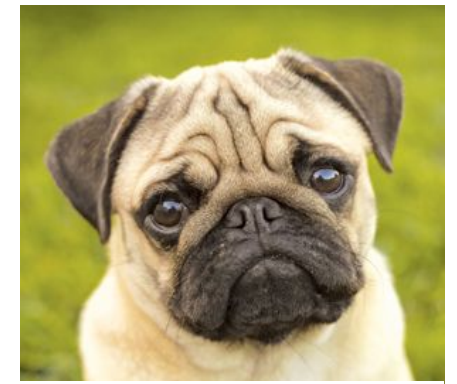
symbolic



Whisky



Tarzan



Snoopy

symbols as representations (e.g. entities, predicates)

easy to integrate domain knowledge: add rules

- powerful logic reasoning tools (prolog)

**“lifted”  
formulation**

```
is_cat(Tarzan).
```

```
% Tarzan is a cat
```

```
is_animal(X) :- is_cat(X)
```

```
% rule: all cats are animals
```

```
?- is_animal(Tarzan)
```

```
% is Tarzan an animal?
```

```
yes
```

- No notion of similarity (Whisky  $\neq$  Tarzan)

```
?- is_animal(Whisky)
```

```
no idea :(
```

**not suited for approximate  
inference**

# Guiding Bootstrapped Relation Extractors

