UMass at English Slot Filling and Cold Start

Unified KB Construction with Universal Schema

Sameer Singh, Limin Yao, David Belanger, Ari Kobren,
Sam Anzaroot, Mike Wick, Alexandre Passos,
Harshal Pandya, Jinho Choi, Brian Martin,
Andrew McCallum

Overview of Relation Extraction

Supervised Classification

Map entity pairs and their mentions to relations

Supervised Classification

Map entity pairs and their mentions to relations

employee(**X**,**Y**)

Supervised Classification

Map entity pairs and their mentions to relations

employee(**X**,**Y**)

Petrie,
UCL

Supervised Classification

Map entity pairs and their mentions to relations

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

KBP

employee(**X**,**Y**)

Petrie, UCL

Supervised Classification

Map entity pairs and their mentions to relations

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

KBP

	employee(**X**,**Y**)
Petrie, UCL	**1**

Supervised Classification

Map entity pairs and their mentions to relations

KBP

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1

Petrie, UCL

Train

Test

Supervised Classification

Map entity pairs and their mentions to relations

KBP

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1

Petrie, UCL

Train

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

Ferguson, Harvard

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

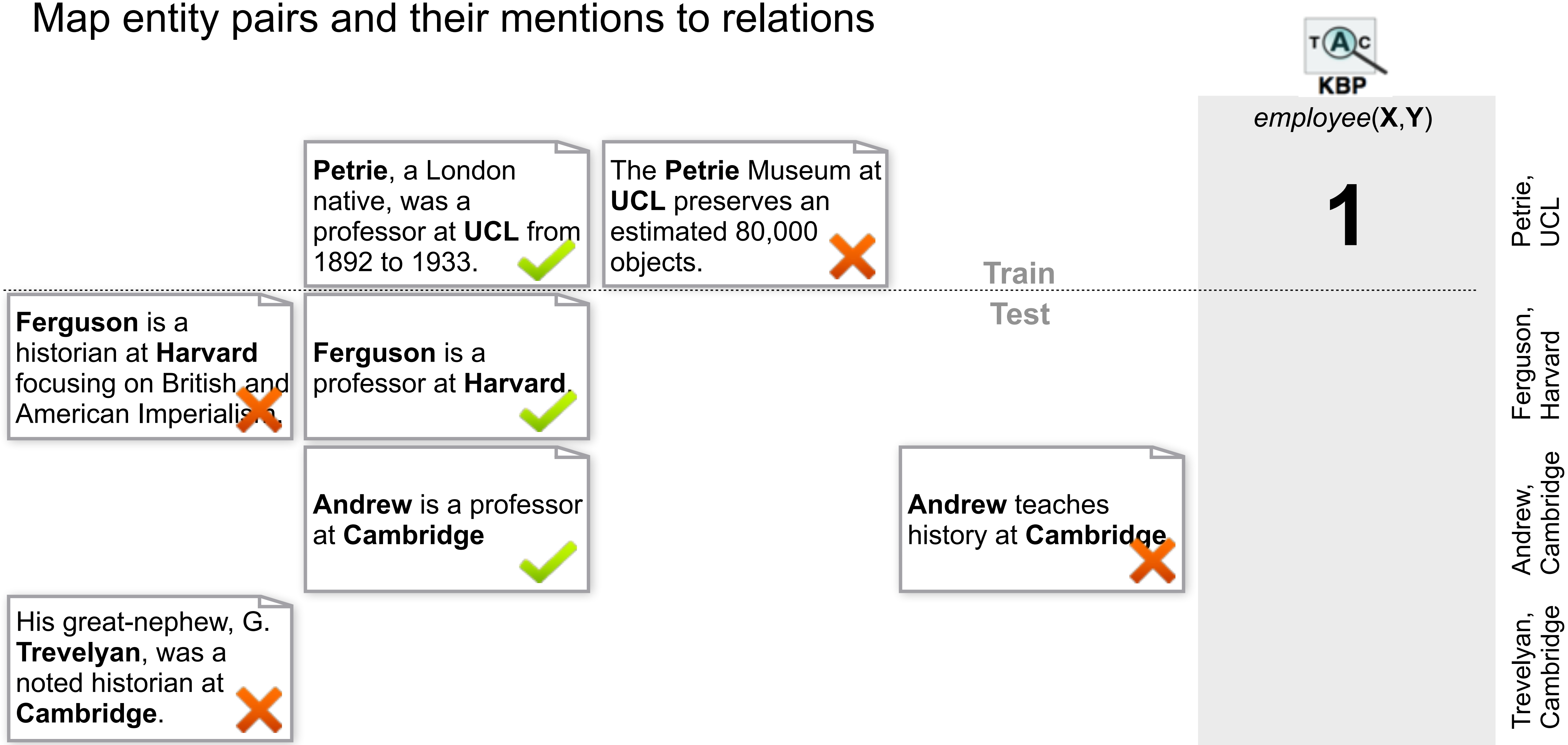
Andrew, Cambridge

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

Trevelyan, Cambridge

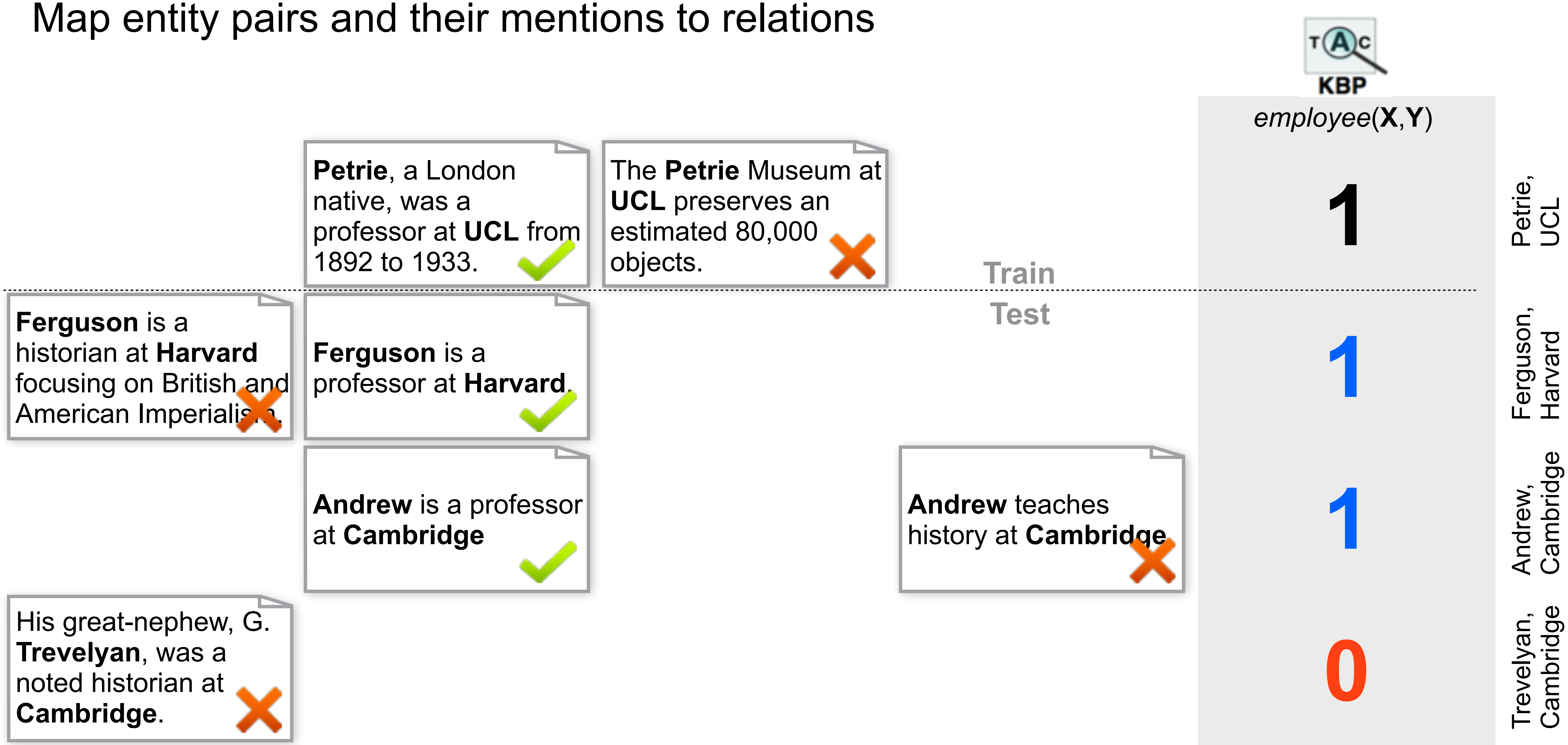
Supervised Classification

Map entity pairs and their mentions to relations



Supervised Classification

Map entity pairs and their mentions to relations



Supervised Classification

Fails for **unseen** patterns

KBP

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Train

1 — Petrie, UCL

Test

Ferguson is a historian at **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

1 — Ferguson, Harvard

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

1 — Andrew, Cambridge

His great-nephew, G. **Trevelyan**, was a noted historian at **Cambridge**.

0 — Trevelyan, Cambridge

Matrix Formulation

Convert problem into a **matrix formulation**

KBP

employee(**X**,**Y**)

Petrie, a London native, was a professor at **UCL** from 1892 to 1933.

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

1

Ferguson *is a historian at* **Harvard** focusing on British and American Imperialism.

Ferguson is a professor at **Harvard**.

1

Andrew is a professor at **Cambridge**

Andrew teaches history at **Cambridge**

1

His great-nephew, G. **Trevelyan**, *was* a noted *historian at* **Cambridge**.

0

Matrix Formulation

Columns correspond to **patterns** between mentions

X-*is-historian-at*-**Y**				*employee*(**X**,**Y**)
	Petrie, a London native, was a professor at **UCL** from 1892 to 1933.	The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.		1
1	**Ferguson** is a professor at **Harvard**.			1
	Andrew is a professor at **Cambridge**		**Andrew** teaches history at **Cambridge**	1
1				0

Matrix Formulation

Columns correspond to **patterns** between mentions

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**
	1
1	1
	1
1	

The **Petrie** Museum at **UCL** preserves an estimated 80,000 objects.

Andrew teaches history at **Cambridge**

KBP

employee(**X**,**Y**)
1
1
1
0

Matrix Formulation

Columns correspond to **patterns** between mentions

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			1
	1		1	1
1				0

Supervised Classification

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			?
	1		1	?
1				?

Train

Test

Supervised Classification

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			?
	1		1	?
1				?

Train

Test

Supervised Classification

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	**1**	Train	**1**
1	**1**		Test	**?**
	1		**1**	**?**
1				**?**

Supervised Classification

X-*is-historian-at*-Y	X-*is-professor-at*-Y	X-*museum-at*-Y	X-*teaches-history-at*-Y	*employee*(X,Y)
	1	1	Train	1
1	1		Test	?
	1		1	?
1				?

does not generalize, costly to label data

Bootstrapping

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1		1
1	1			?
	1		1	?
1				?

Train

Test

Bootstrapping

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1	Train	1
1	1		Test	?
	1		1	?
1				?

Bootstrapping

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)
	1	1	Train	1
1	1		Test	?
	1		1	?
1				?

difficult to control generalization, still relies on labeled data

Distant Supervision

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)	/employee
	1			1 Train	1
1	1			Test ?	
	1		1	?	
1				?	

Distant Supervision

TAC KBP | Freebase

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)	/employee
	1			1 Train	1
1	1			Test ?	
	1		1	?	
1				?	

Distant Supervision

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)	/employee
	1	**1**		**1** Train	**1**
1	**1**			Test **?**	
	1		**1**	**?**	
1				**?**	

Distant Supervision

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)	/employee
	1	1		1 Train	1
1	1			Test ?	
	1		1	?	
1				?	

Distant Supervision

X-*is-historian-at*-**Y**	**X**-*is-professor-at*-**Y**	**X**-*museum-at*-**Y**	**X**-*teaches-history-at*-**Y**	*employee*(**X**,**Y**)	/employee
	1	$\boxed{1}$		1 Train	1
1	1			Test ?	
	1		1	?	
1				?	

need alignment to external schema, assumptions may be inaccurate

Other Problems

Other Problems

- A relation may be implied by **multiple extractions**
  - cityOfHQ(X,Y): many weak co-occurrences of X & Y

Other Problems

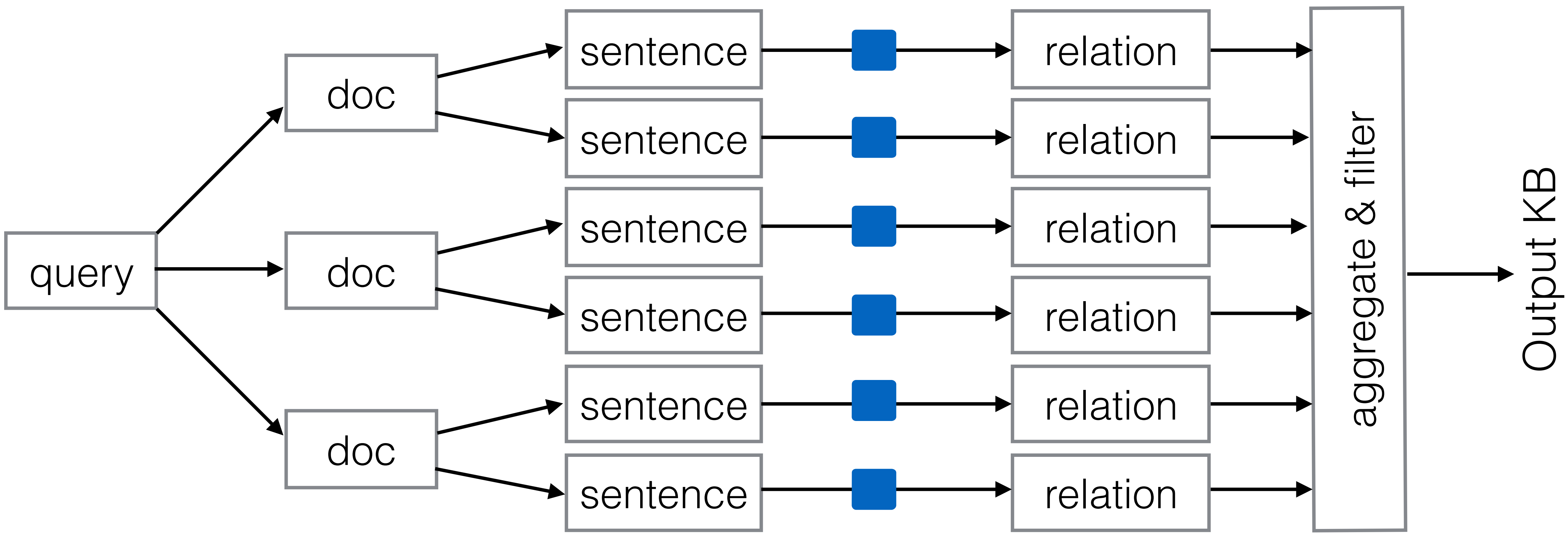
- A relation may be implied by **multiple extractions**
  - cityOfHQ(X,Y): many weak co-occurrences of X & Y
- A relation may be implied by **other predicted relations**
  - topEmployeeOf → employeeOf
  - cityOfResidence → cityOfDeath (probabilistic!)

Other Problems

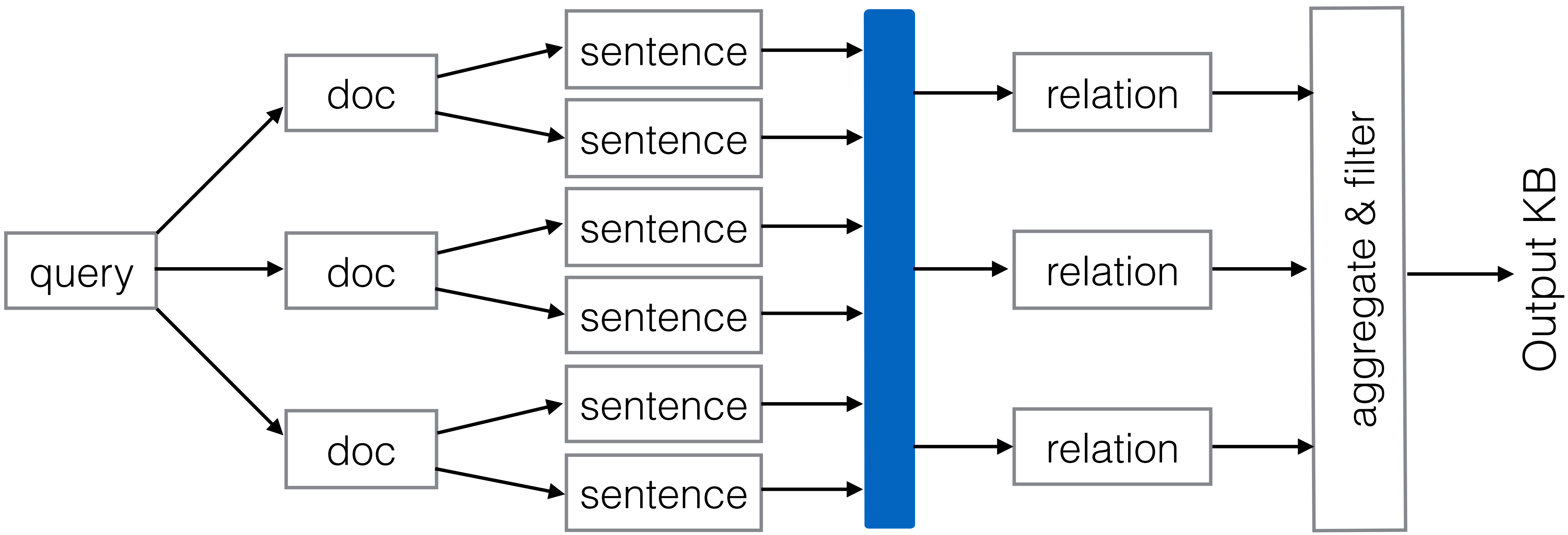
- A relation may be implied by **multiple extractions**
  - cityOfHQ(X,Y): many weak co-occurrences of X & Y
- A relation may be implied by **other predicted relations**
  - topEmployeeOf → employeeOf
  - cityOfResidence → cityOfDeath (probabilistic!)
- A relation may require **context from multiple entities** (that may not be query entities)
  - *X was in Y*: school, company, city, state, country?
  - *X, president of Y*: headOfState, topEmployeeOf

Unified KB Construction

Query-time Pipeline

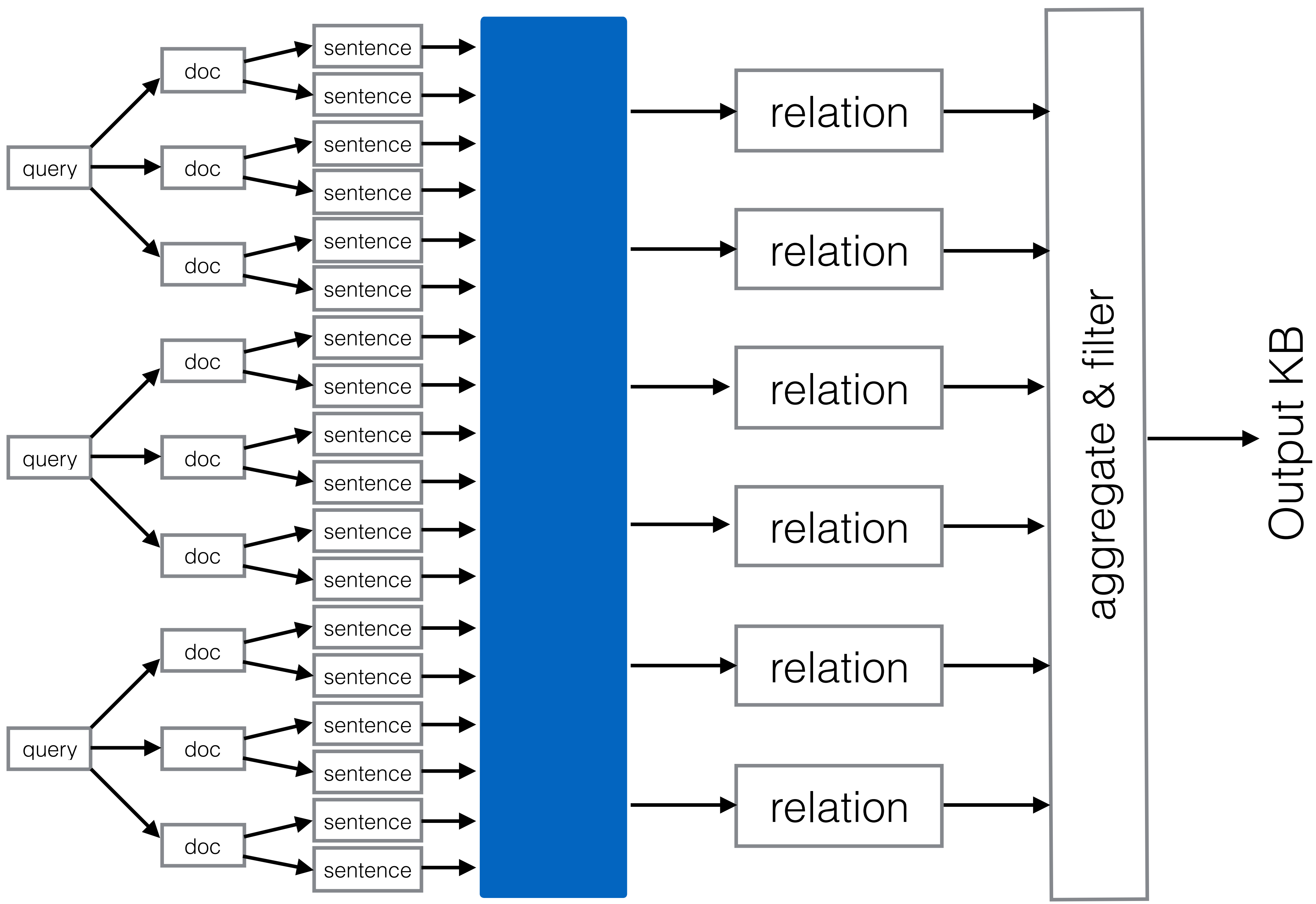


Query-time Pipeline

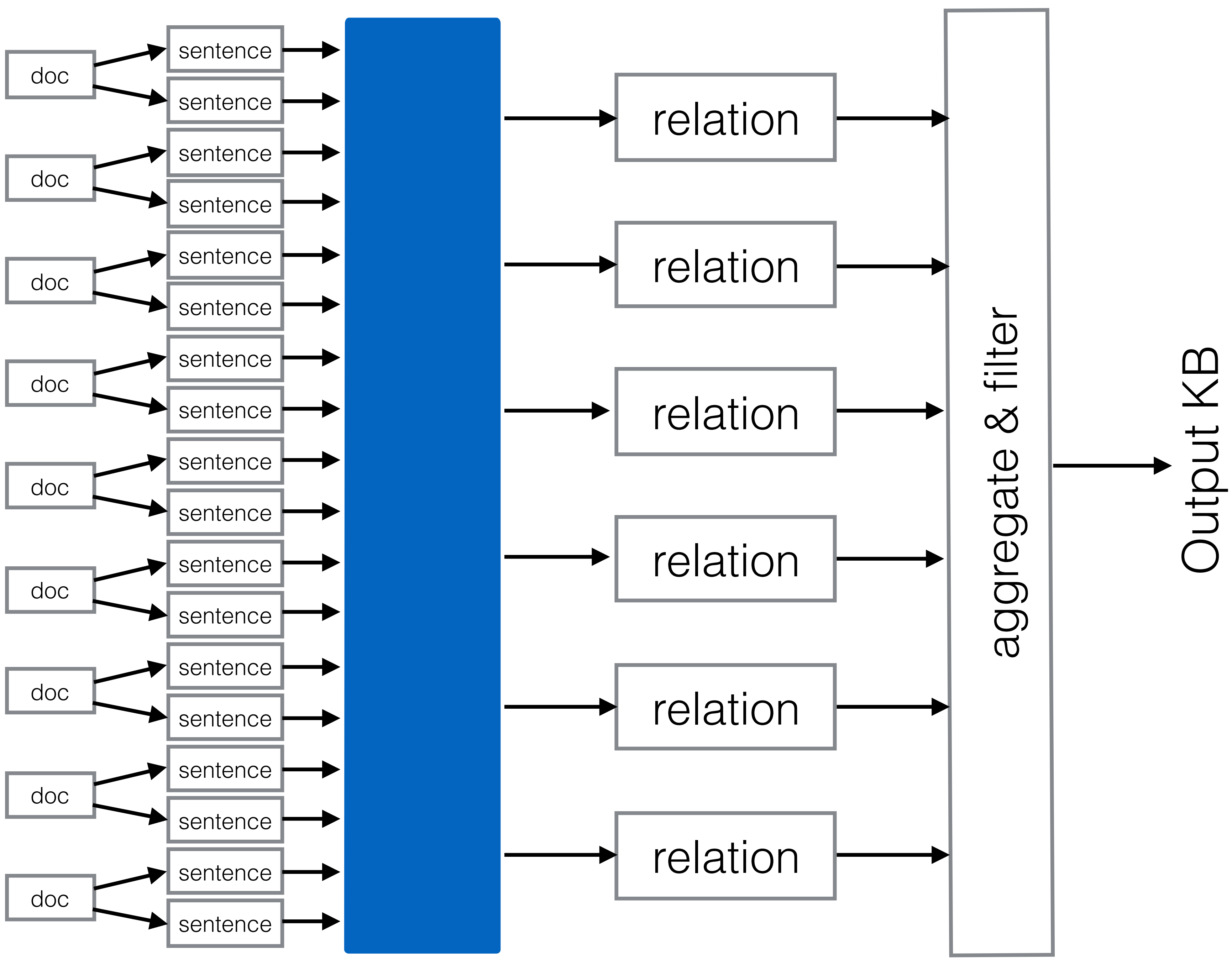


Similar to recent distant supervision systems.

Model all the queries



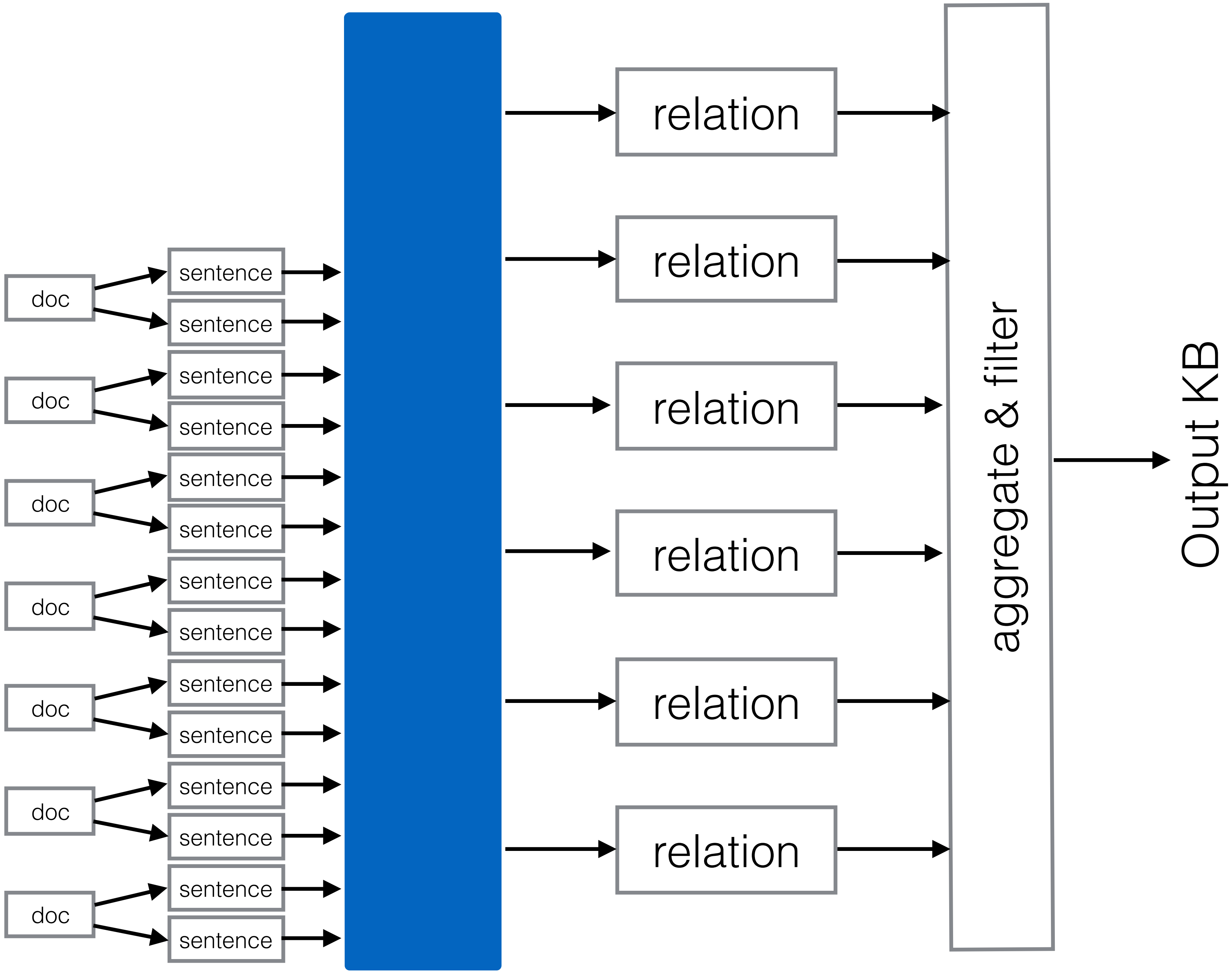
Model all the entities!



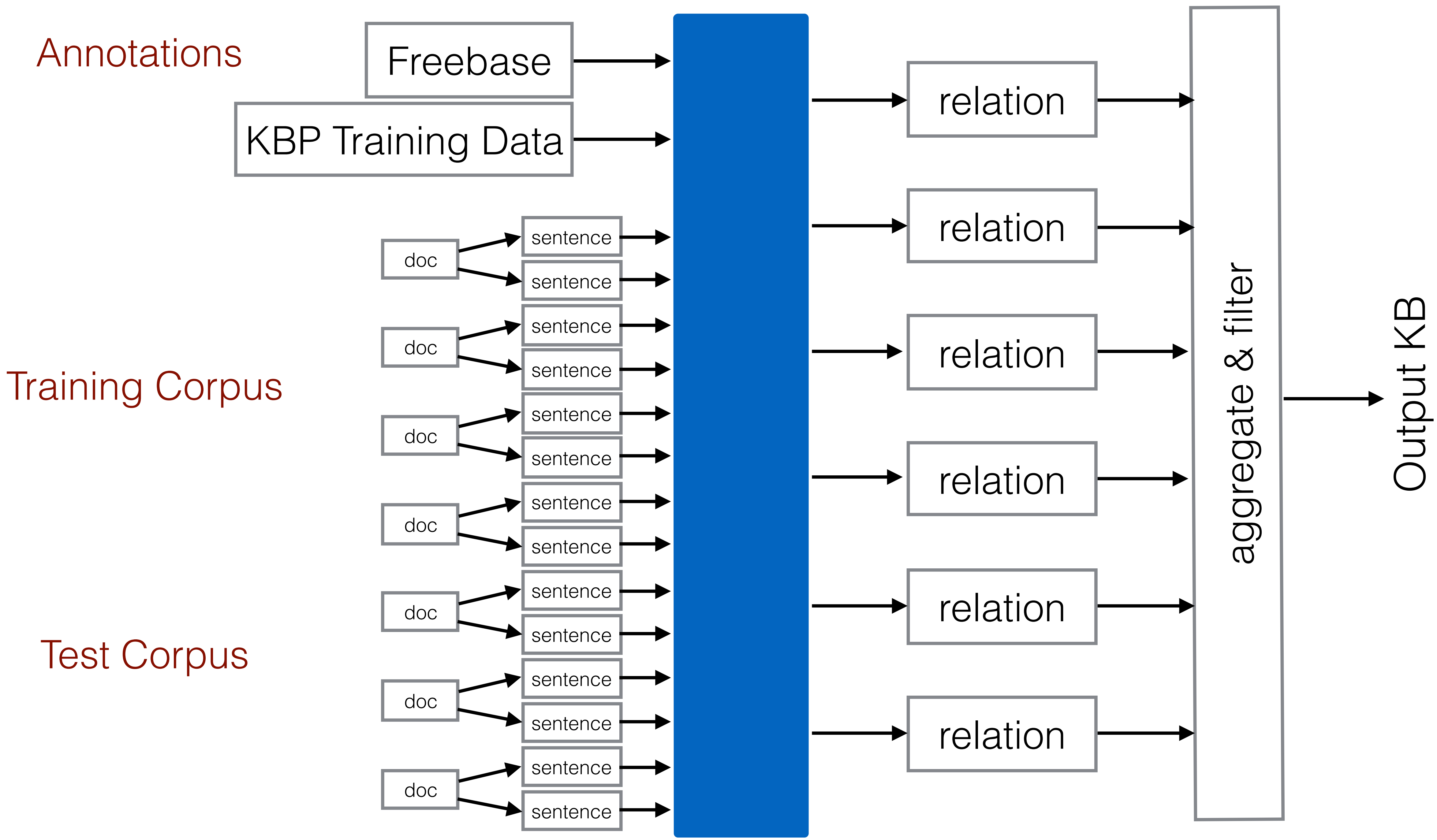
… and all forms of "evidence"

Training Corpus

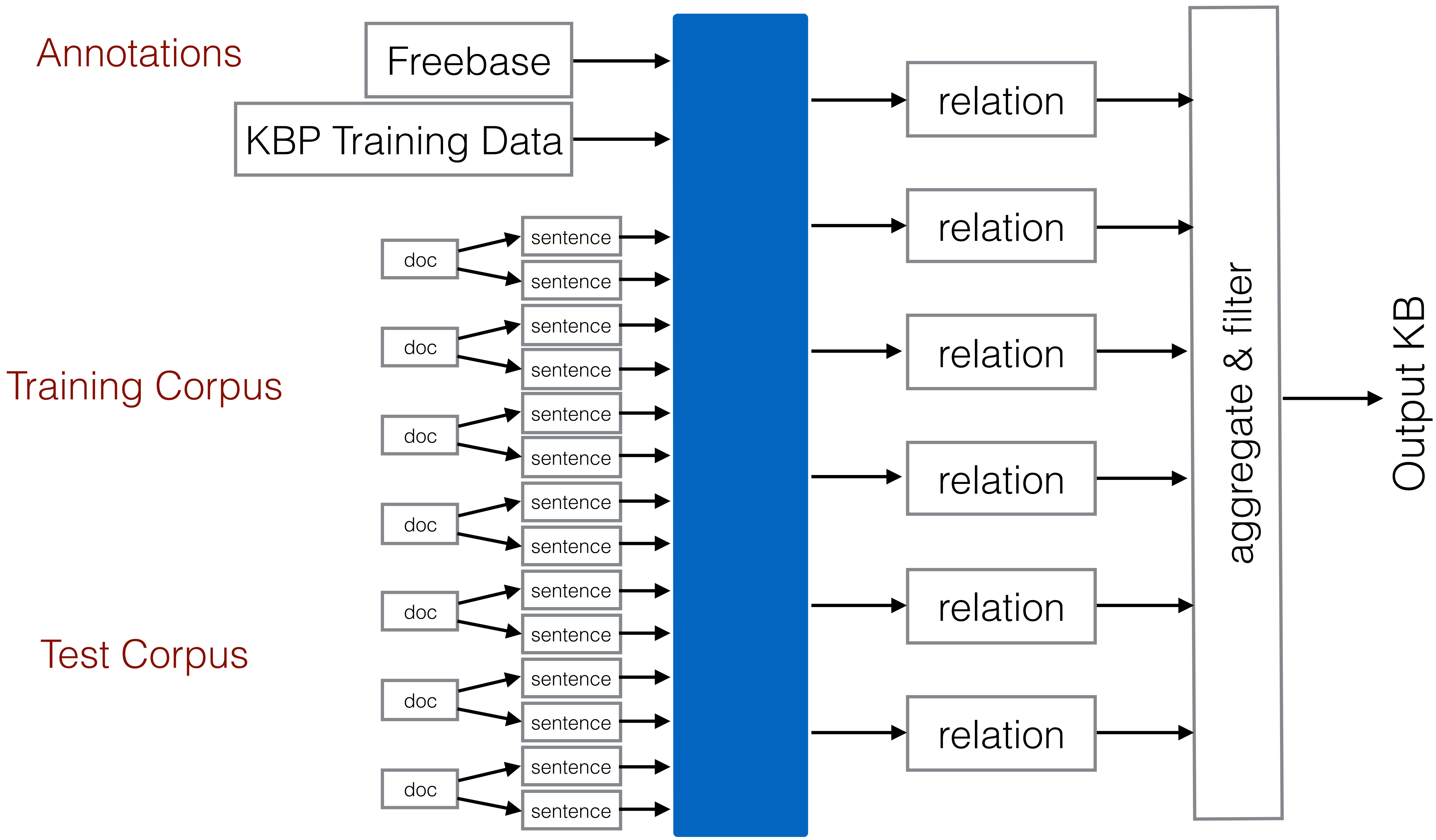
Test Corpus



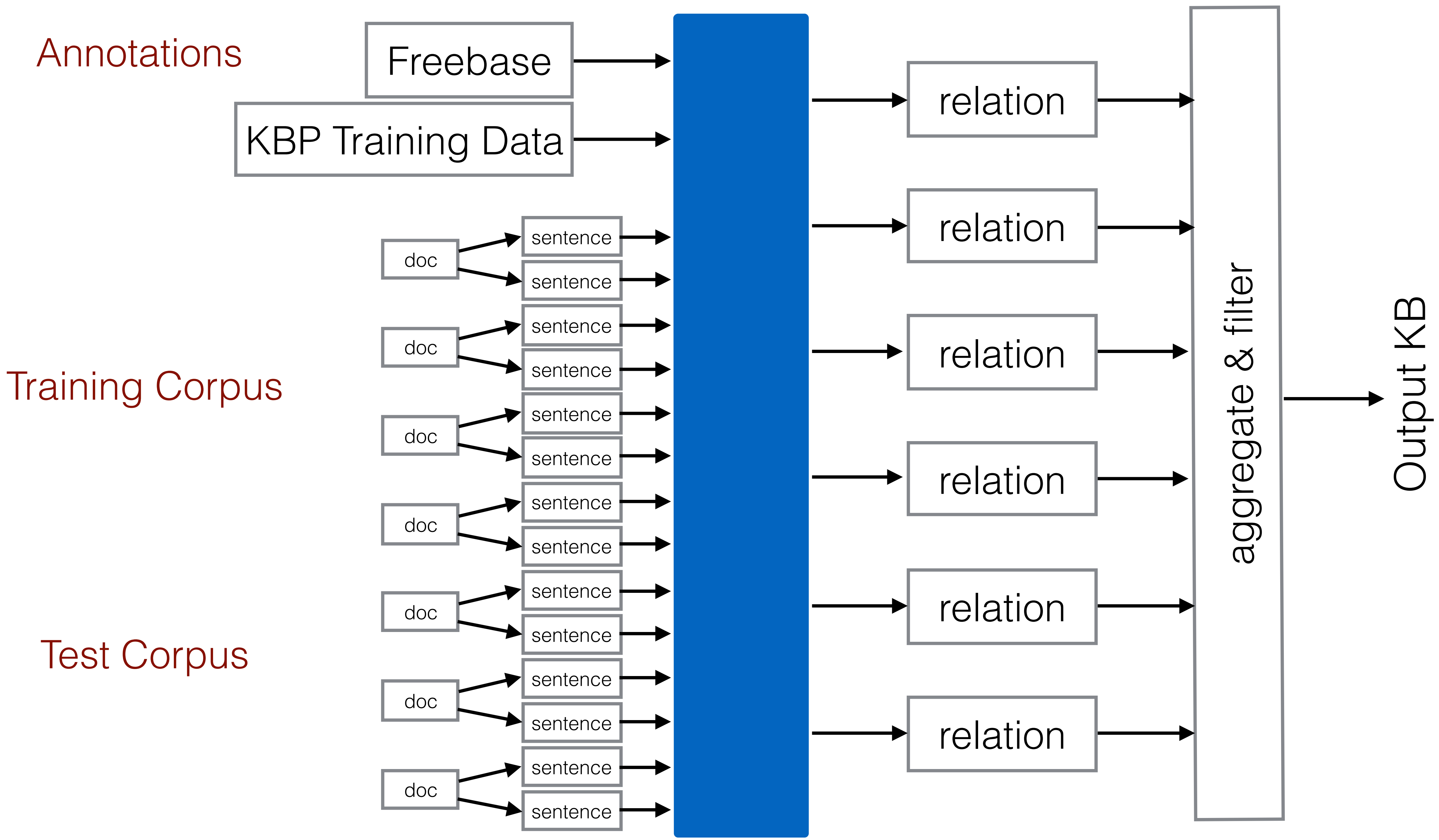
… and all forms of “evidence”



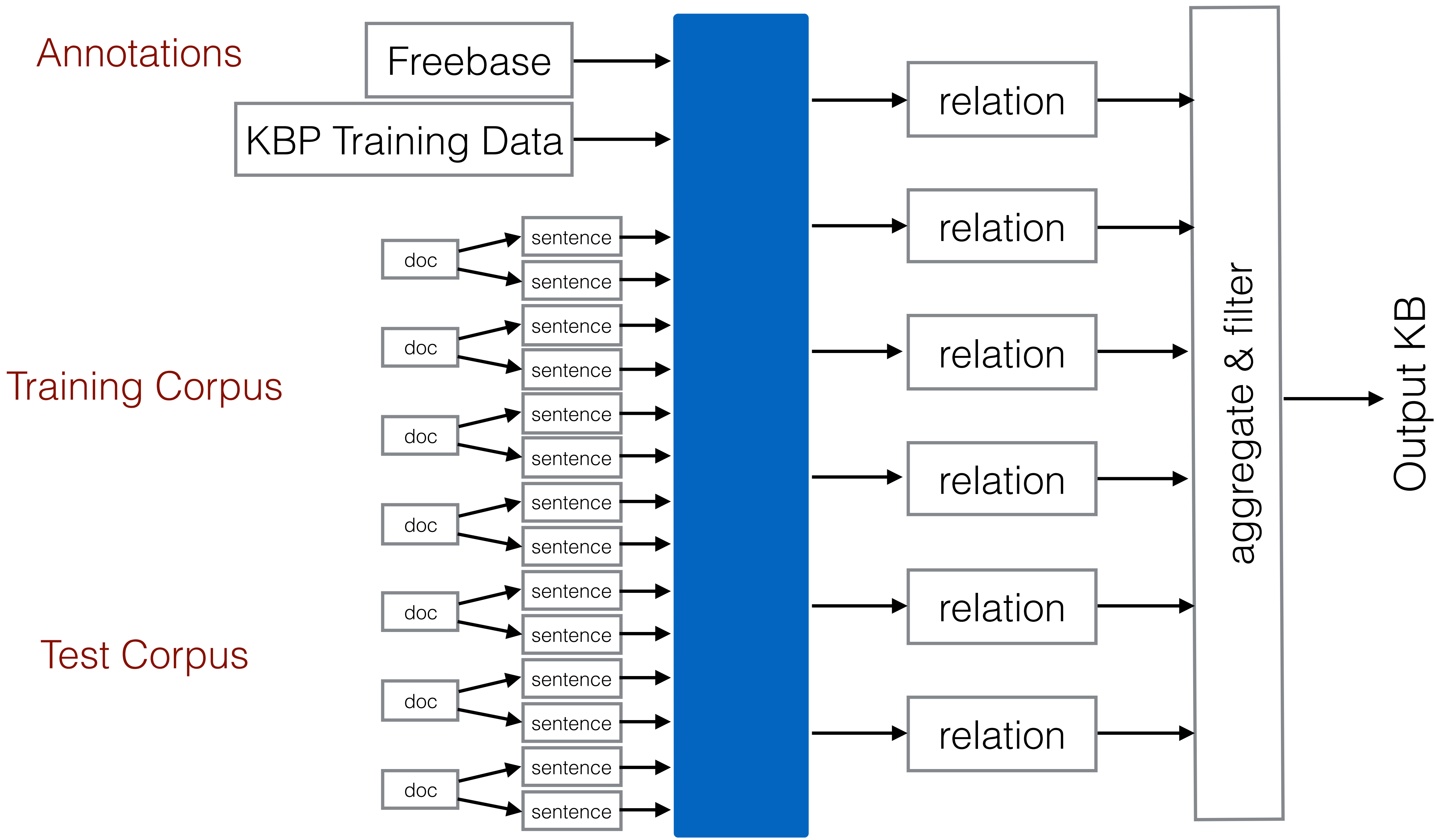
… and all forms of "evidence"



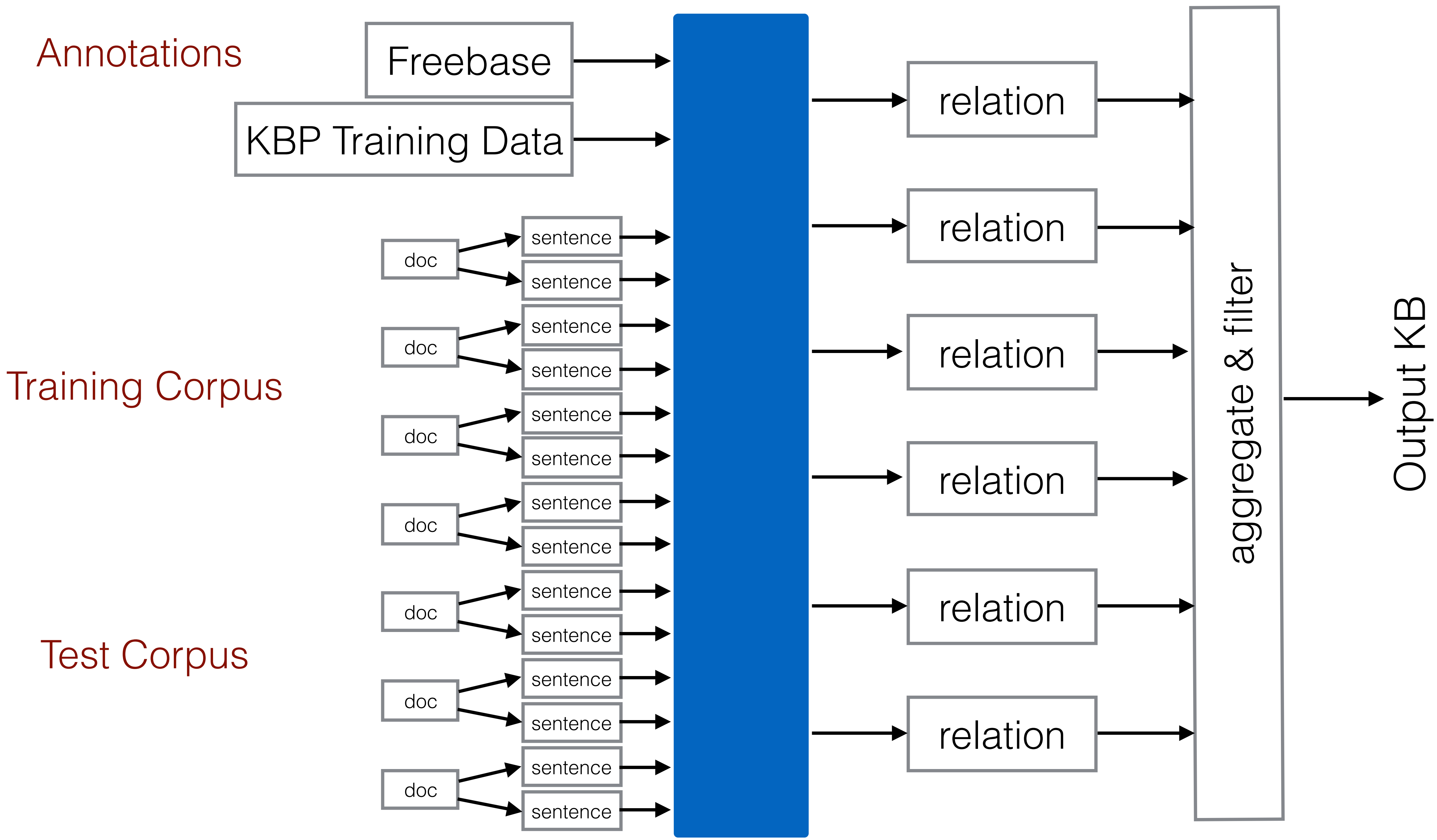
… and all forms of "evidence"



… and all forms of “evidence”



… and all forms of "evidence"



Challenges

Challenges

- Needs to accurately learn all **correlations** amongst:
  - KBP and Freebase relations
  - patterns and relations
  - entity context and relations

Challenges

- Needs to accurately learn all **correlations** amongst:
  - KBP and Freebase relations
  - patterns and relations
  - entity context and relations
- Needs to be **scalable**, since we will run it over:
  - all documents in the corpus
  - all entities that appear in them
  - all facts from the external KB and TAC RefKB

Universal Schema

Universal Schema

Text documents: relations from dependency parses

	president of	*prime minister of*	*chancellor of*	*chief executive*	*leader of*	*head of state*	headOf	Top Member
Obama, U.S.								
Merkel, Germany								
S Harper Canada								
V Putin Russia								
Larry Page Google								
V. Rometty IBM								
Tim Cook Apple								
E Grimson MIT								

330k columns

3 million rows

[Yao, Riedel, McCallum, AKBC 2012]

Universal Schema

Text documents: relations from dependency parses

Freebase

TAC KBP

	president of	*prime minister of*	*chancellor of*	*chief executive*	*leader of*	*head of state*	headOf	Top Member
Obama, U.S.	1			1			1	
Merkel, Germany			1		1	1	1	
S Harper Canada		1			1		1	
V Putin Russia	1				1	1	1	
Larry Page Google				1			1	1
V. Rometty IBM	1			1	1			1
Tim Cook Apple				1			1	1
E Grimson MIT			1				1	

330k columns

3 million rows

Combination of structured and OpenIE

[Yao, Riedel, McCallum, AKBC 2012]

Universal Schema

Text documents: relations from dependency parses

	president of	*prime minister of*	*chancellor of*	*chief executive*	*leader of*	*head of state*	headOf	Top Member
Obama, U.S.	1			1			1	
Merkel, Germany			1		1	1	1	
S Harper Canada		1			1		1	
V Putin Russia	1				1	1	1	
Larry Page Google				1			1	1
V. Rometty IBM	1			1	1			1
Tim Cook Apple				1			1	1
E Grimson MIT			1				1	

330k columns

3 million rows

Model & fill in matrix with *Low-Rank Matrix Factorization* (ala NetFlix)

[Yao, Riedel, McCallum, AKBC 2012]

Universal Schema

Text documents: relations from dependency parses

Freebase

TAC KBP

	president of	*prime minister of*	*chancellor of*	*chief executive*	*leader of*	*head of state*	headOf	Top Member
Obama, U.S.	1			1			1	
Merkel, Germany			1		1	1	1	
S Harper Canada		1			1		1	
V Putin Russia	1				1	1	1	
Larry Page Google				1			1	1
V. Rometty IBM	1			1	1			1
Tim Cook Apple				1			1	1
E Grimson MIT			1				1	

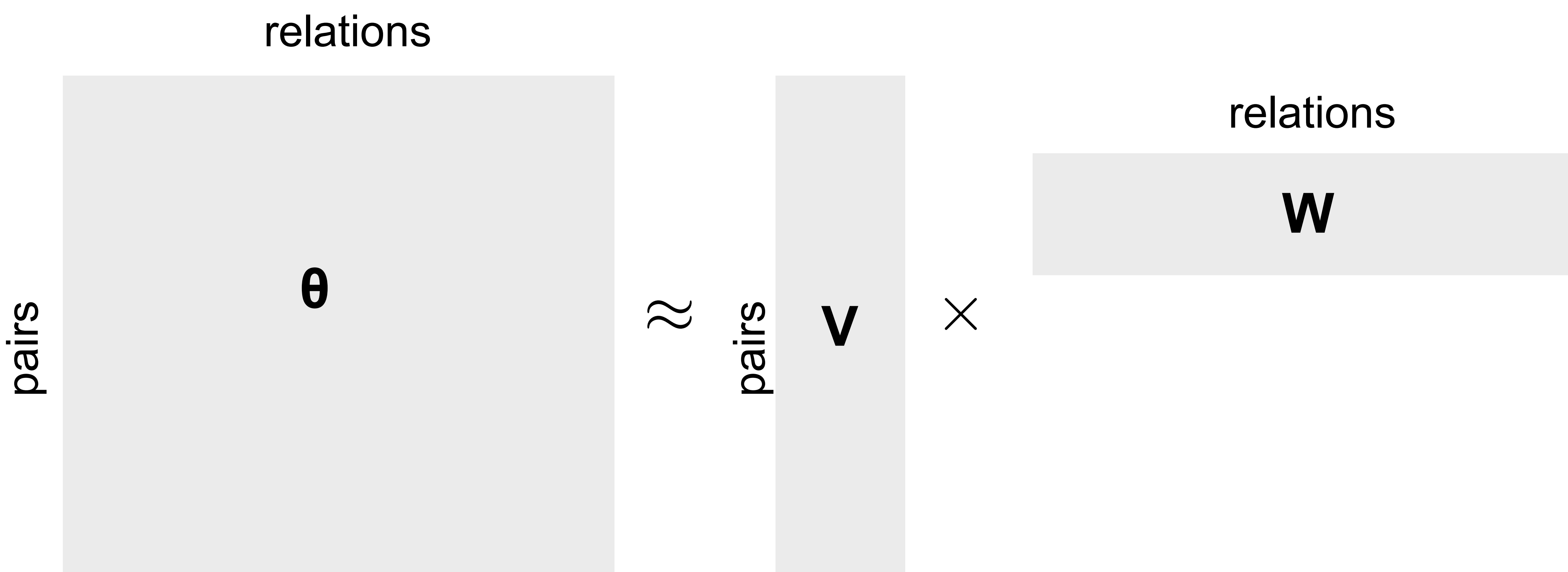
330k columns

3 million rows

Model & fill in matrix with *Low-Rank Matrix Factorization* (ala NetFlix)

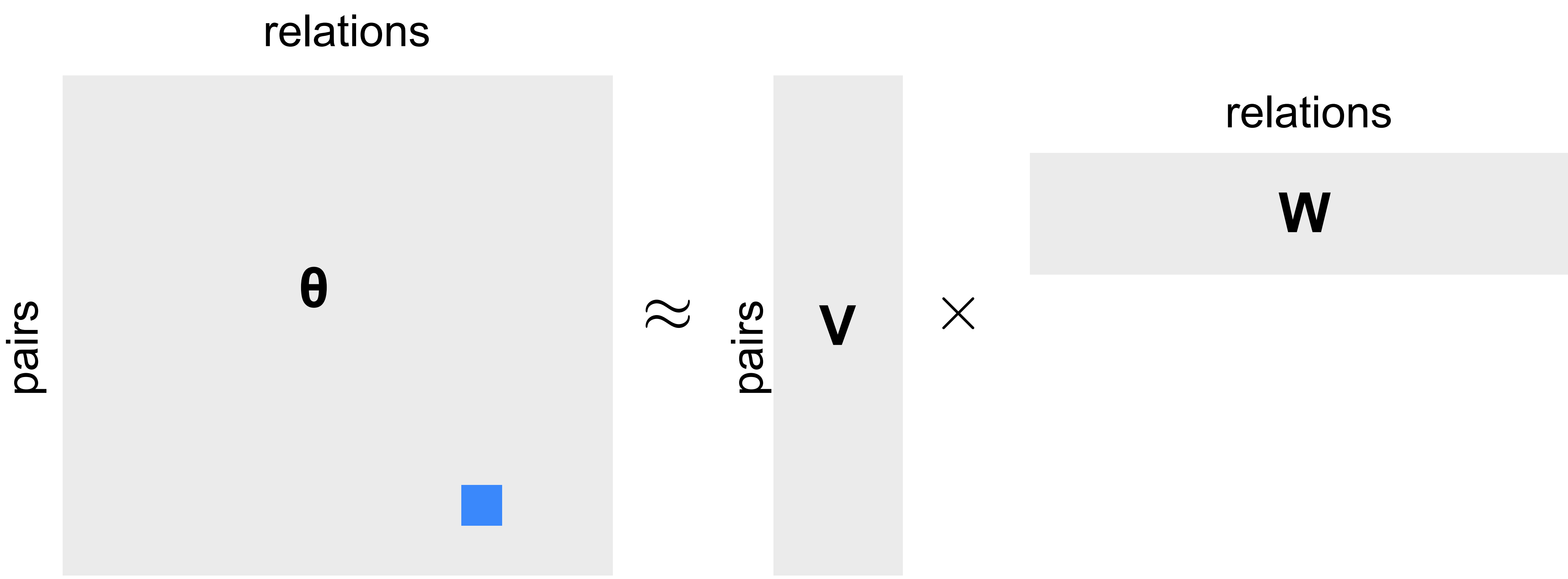
[Yao, Riedel, McCallum, AKBC 2012]

Model



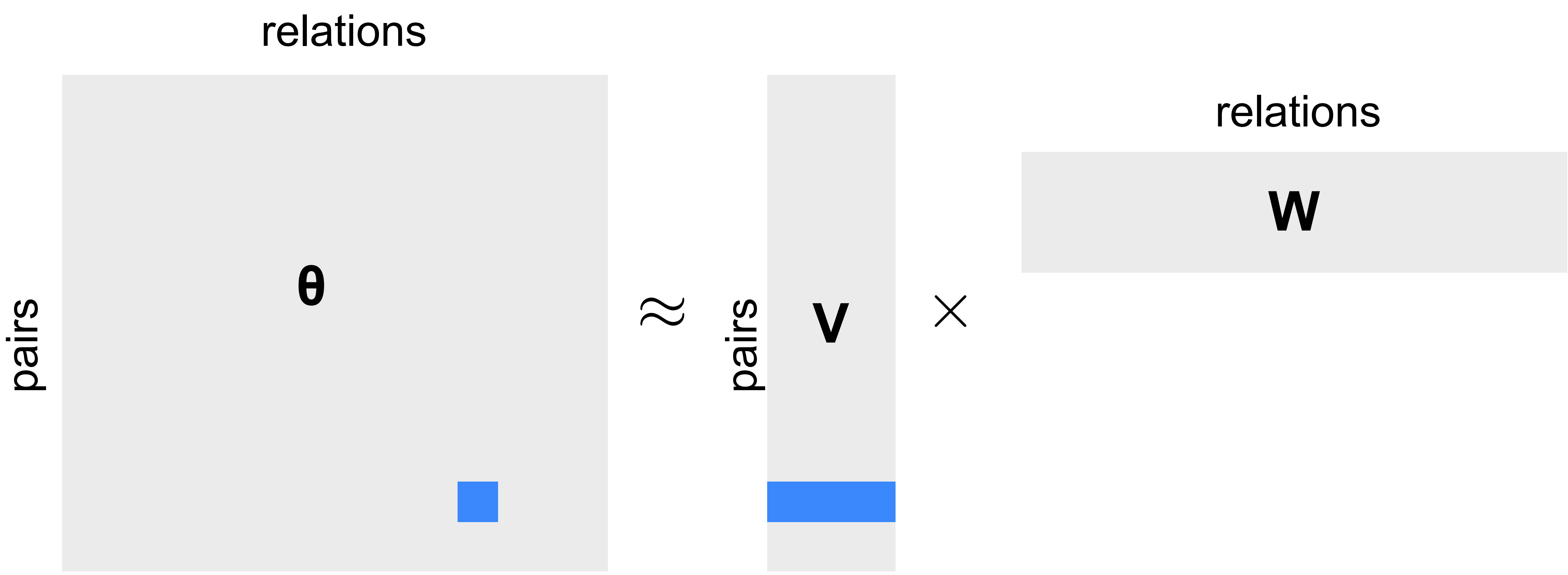
$$\theta^{x,y}_{\text{emp}} =< \mathbf{v}^{x,y}, \mathbf{w}_{\text{emp}} > \qquad p(y^{x,y}_{\text{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\text{emp}}$$

Model



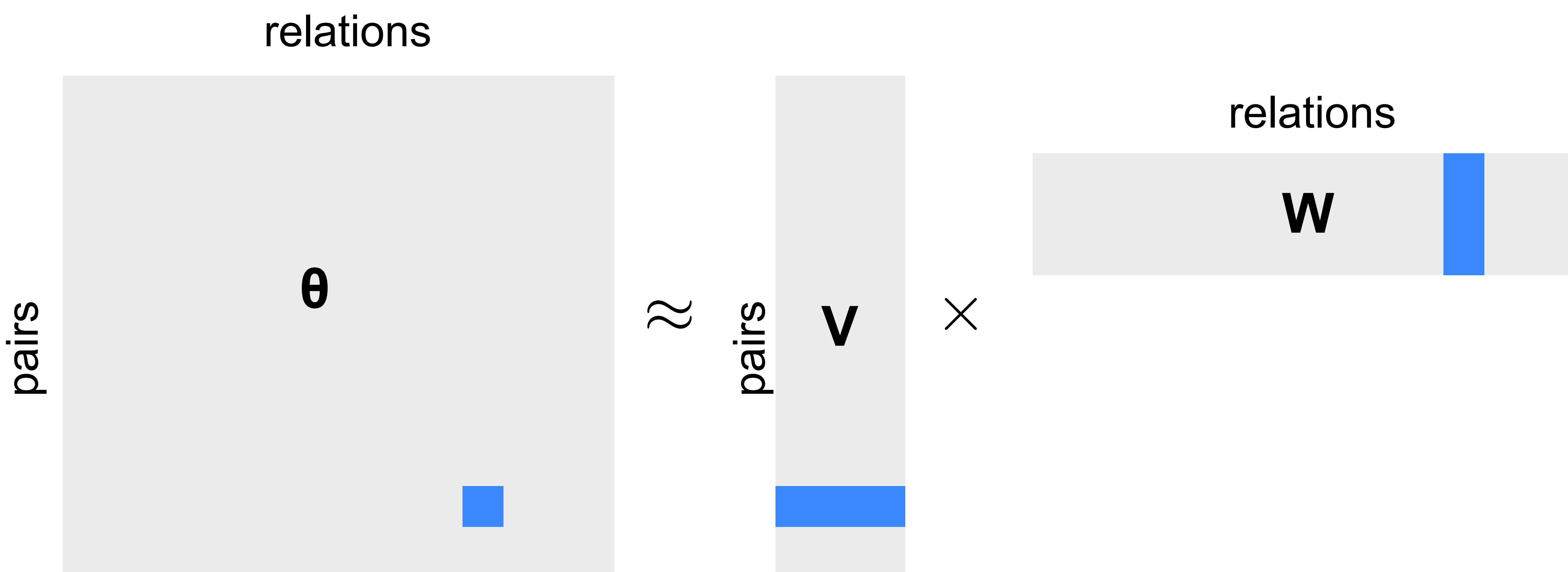
$$\theta^{x,y}_{\mathrm{emp}} = <\mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{emp}}> \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\mathrm{emp}}$$

Model



$$\theta^{x,y}_{\text{emp}} =< \mathbf{v}^{x,y}, \mathbf{w}_{\text{emp}} > \qquad p(y^{x,y}_{\text{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\text{emp}}$$

Model



$$\theta^{x,y}_{\mathrm{emp}} =< \mathbf{v}^{x,y}, \mathbf{w}_{\mathrm{emp}} > \qquad p(y^{x,y}_{\mathrm{emp}} = 1 | \ldots) \propto \exp \theta^{x,y}_{\mathrm{emp}}$$

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:
    - Do not instantiate the whole matrix!

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:
  - Do not instantiate the whole matrix!
  - Do not hold all the observed cells in memory

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:
  - Do not instantiate the whole matrix!
  - Do not hold all the observed cells in memory
  - Each iteration linear in the no. of observations

Training

- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:
  - Do not instantiate the whole matrix!
  - Do not hold all the observed cells in memory
  - Each iteration linear in the no. of observations
- **Solution**: Stochastic Gradient Descent!

Training

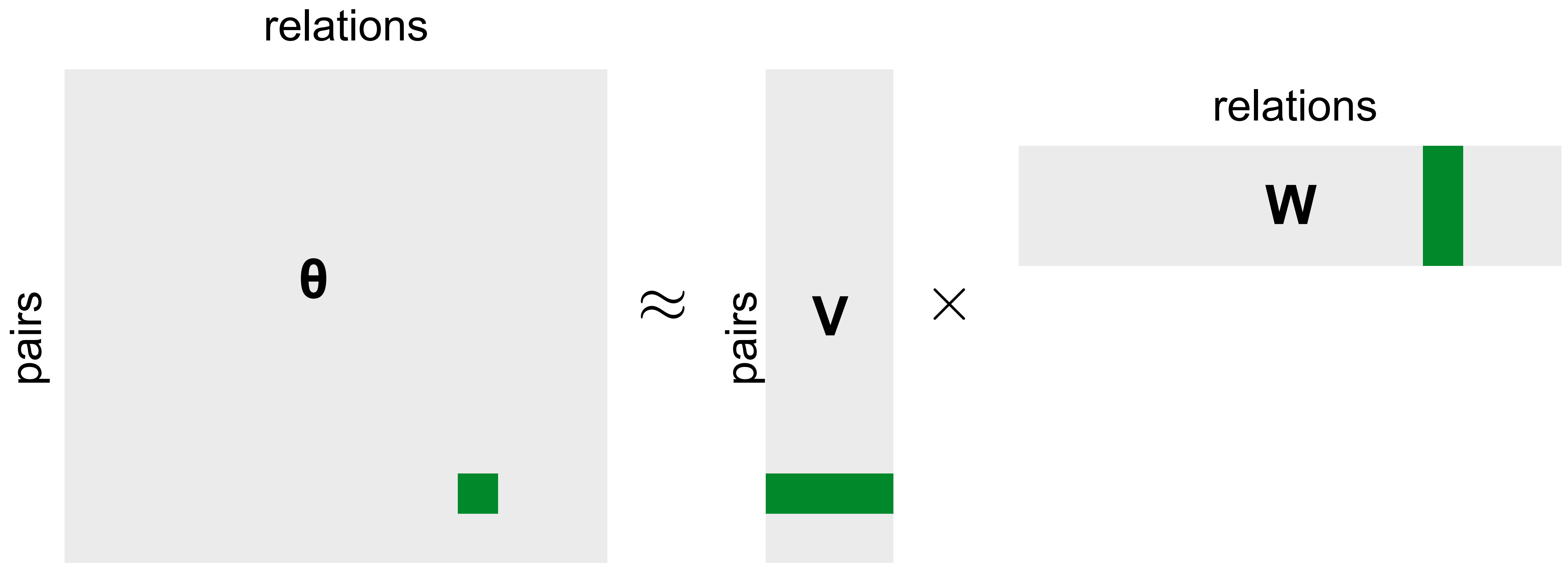
- **Loss Function**: $\max_{\mathbf{v},\mathbf{w}} \log \prod_{x,y,r} \exp\langle \mathbf{v}^{x,y}, \mathbf{w}_r \rangle - \lambda(||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2)$
- Desiderata from the training algorithm:
  - Do not instantiate the whole matrix!
  - Do not hold all the observed cells in memory
  - Each iteration linear in the no. of observations
- **Solution**: Stochastic Gradient Descent!

Stochastic Gradient Descent

Stochastic Gradient Descent

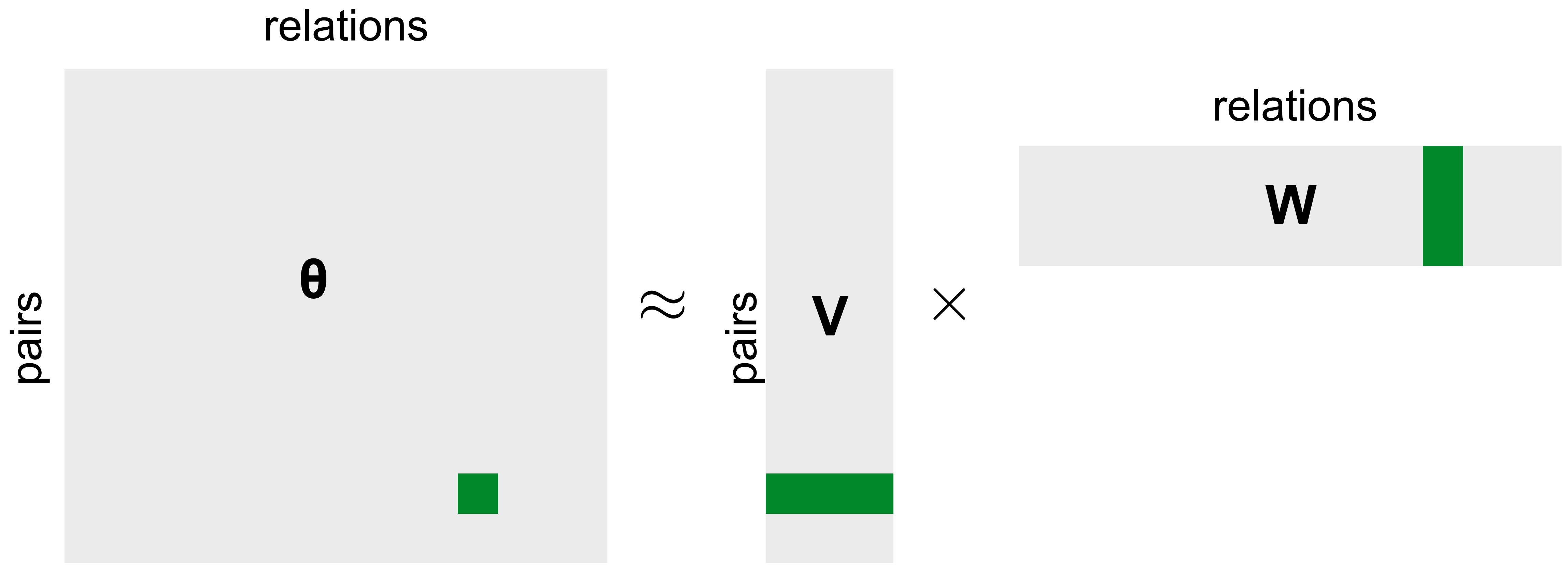
- Pick an observed cell, $\theta^{r}_{x,y}$:

Stochastic Gradient Descent



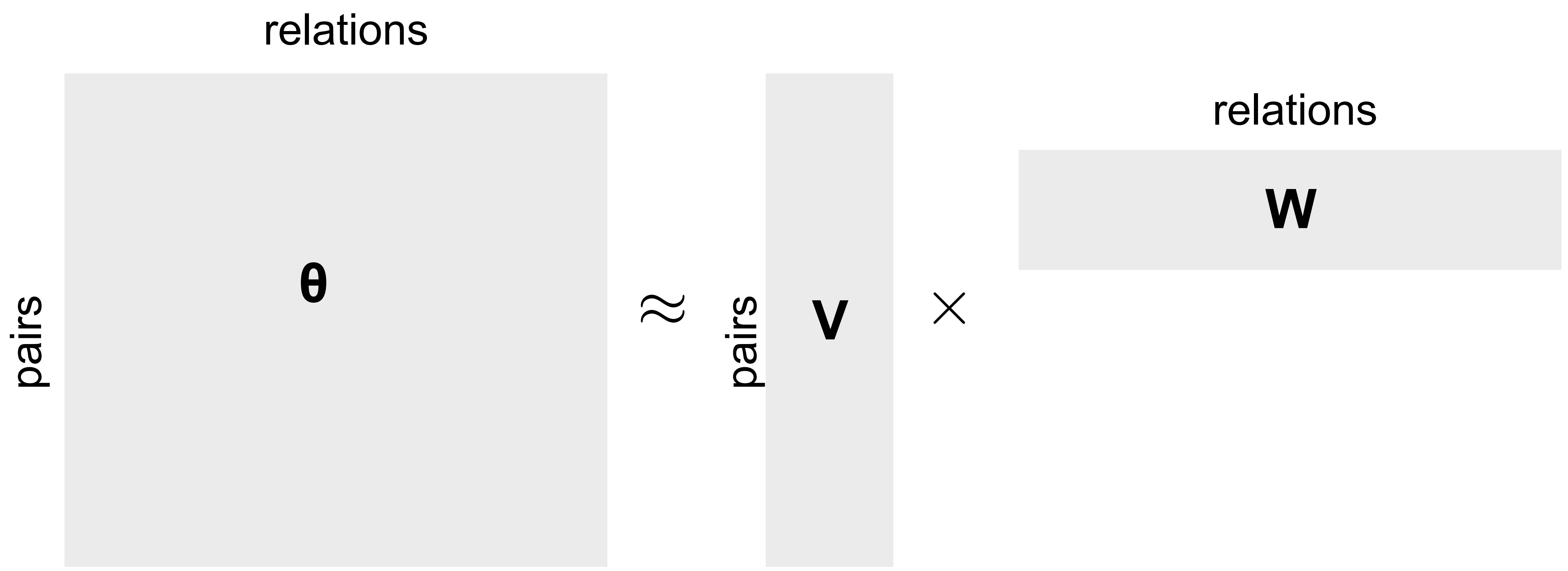
- Pick an observed cell, $\theta^r_{x,y}$:

Stochastic Gradient Descent



- Pick an observed cell, $\theta^r_{x,y}$:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is higher

Stochastic Gradient Descent

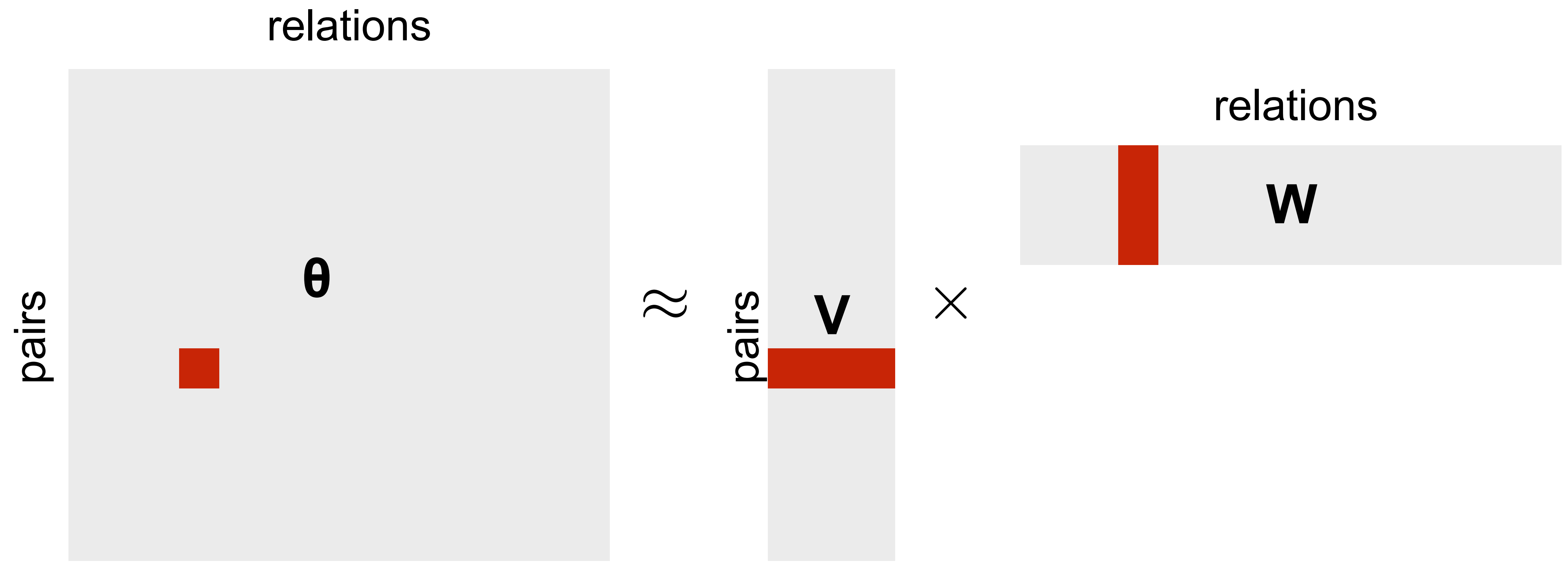


- Pick an observed cell, $\theta^r_{x,y}$:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is higher

Stochastic Gradient Descent

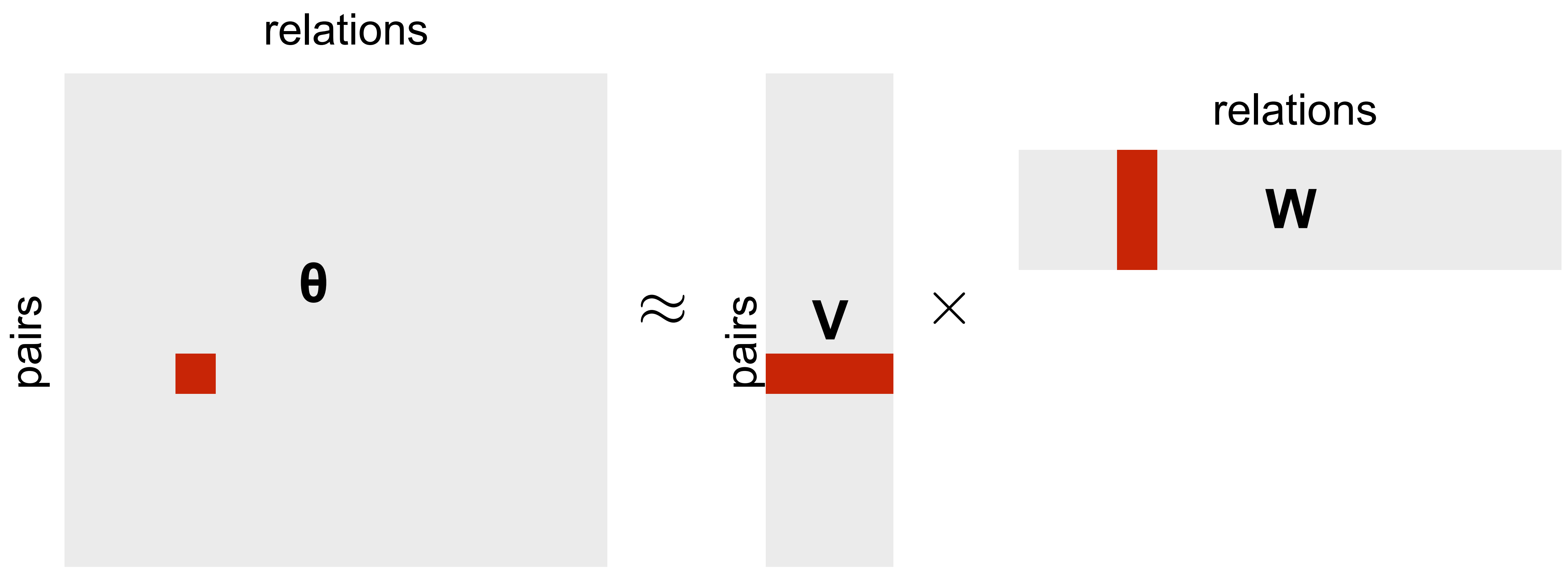
- Pick an observed cell, $\theta^r_{x,y}$:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is higher
- Pick any random cell, assume it is negative:

Stochastic Gradient Descent



- Pick an observed cell, $\theta^r_{x,y}$:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is higher
- Pick any random cell, assume it is negative:

Stochastic Gradient Descent



- Pick an observed cell, $\theta^r_{x,y}$:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is higher
- Pick any random cell, assume it is negative:
  - Update $\mathbf{v}^{x,y}$ & $\mathbf{w}^r$ such that $\theta^r_{x,y}$ is lower

Latent Embeddings

Latent Embeddings

- Relation embeddings, **w**

Latent Embeddings

- Relation embeddings, **w**
    - Similar embedding for 2 relations denote they are **paraphrases**

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) $\subset$ w(employeeOf)

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) ⊂ w(employeeOf)
    - topEmployeeOf(X,Y) → employeeOf(X,Y)

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) ⊂ w(employeeOf)
    - topEmployeeOf(X,Y) → employeeOf(X,Y)
- Entity Pair embeddings, **v**

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) ⊂ w(employeeOf)
    - topEmployeeOf(X,Y) → employeeOf(X,Y)
- Entity Pair embeddings, **v**
  - similar entity pairs denote similar relations between them

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) ⊂ w(employeeOf)
    - topEmployeeOf(X,Y) → employeeOf(X,Y)
- Entity Pair embeddings, **v**
  - similar entity pairs denote similar relations between them
  - entity pairs may describe multiple "**relations**"

Latent Embeddings

- Relation embeddings, **w**
  - Similar embedding for 2 relations denote they are **paraphrases**
    - *is married to*, spouseOf(X,Y), /person/spouse
  - One embedding can be **contained** by another
    - w(topEmployeeOf) $\subset$ w(employeeOf)
    - topEmployeeOf(X,Y) $\rightarrow$ employeeOf(X,Y)
- Entity Pair embeddings, **v**
  - similar entity pairs denote similar relations between them
  - entity pairs may describe multiple "**relations**"
    - independent foundedBy and employeeOf relations

Similar Embeddings

similar underlying embedding

similar embedding

	X **own percentage of** Y	X **buy stake in** Y
Time, Inc Amer. Tel. and Comm.	**1**	**1**
Volvo Scania A.B.		**1**
Campeau Federated Dept Stores		
Apple HP		

Similar Embeddings

similar underlying embedding

similar embedding

	X **own percentage of** Y	X **buy stake in** Y
Time, Inc Amer. Tel. and Comm.	**1**	**1**
Volvo Scania A.B.		**1**
Campeau Federated Dept Stores		
Apple HP		

Successfully predicts
"Volvo owns percentage of Scania A.B."
from "Volvo bought a stake in Scania A.B."

Implications

X historian at Y → X professor at Y

(Freeman,Harvard)
→
(Boyle,OhioState)

	X **professor at** Y	X **historian at** Y
Kevin Boyle Ohio State		**1**
R. Freeman Harvard	**1**	

Implications

X historian at Y → X professor at Y

	X **professor at** Y	X **historian at** Y
Kevin Boyle Ohio State		**1**
R. Freeman Harvard	**1**	

(Freeman,Harvard)
→
(Boyle,OhioState)

Learns asymmetric entailment:

PER historian at UNIV → PER professor at UNIV

but PER professor at UNIV ↛ PER historian at UNIV

Experiments

20 years NYTimes

extract entity mentions, perform entity resolution

350k entity pairs, 23k unique relation surface forms

Freebase

6k entity pairs resolved with NYTimes pairs

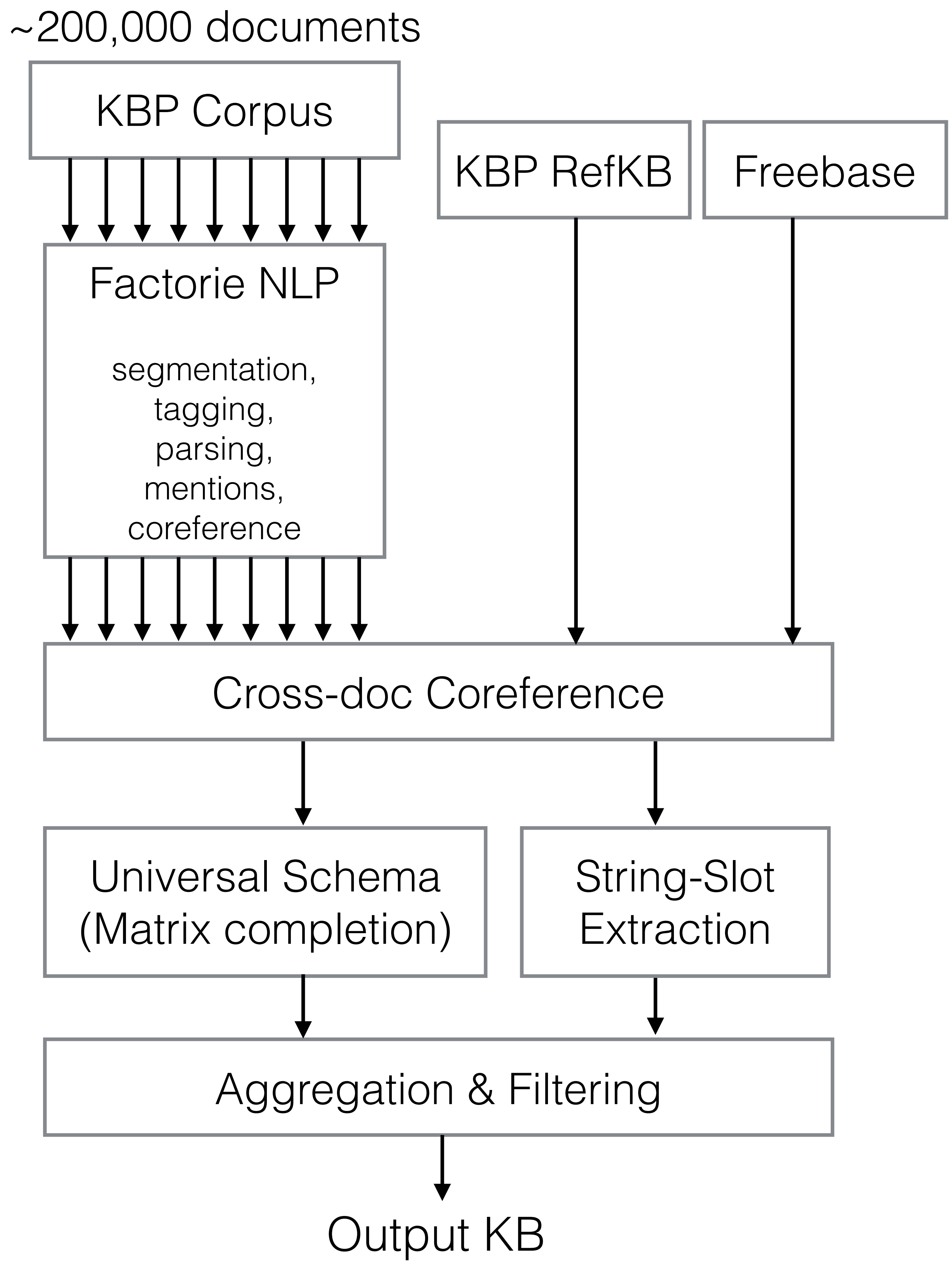
116 relations

[Riedel, Yao, McCallum, Marlin, NAACL 2013]

Relation	Distant Supervision [UMass 2011]	Unsup Clustering [UMass 2012]	MIML > Mintz, Hoffmann [Stanford 2012]	Univ Schema [UMass 2013]
place lived	0.18	0.28	0.1	**0.59**
work for	0.67	0.64	0.7	**0.75**
contained by	0.48	0.51	0.54	**0.68**
author/work	0.5	0.51	0.52	**0.61**
nationality	0.14	**0.4**	0.13	0.19
parent comp	0.14	0.25	0.62	**0.76**
death	0.79	0.79	**0.86**	0.83
birth	0.78	0.75	0.82	**0.83**
person parent	0.24	0.27	**0.58**	0.53
Weighted Average (on even more types)	0.48	0.52	0.57	**0.69**

[Riedel, Yao, McCallum, Marlin, NAACL 2013]

KBP Pipeline


~200,000 documents
KBP Corpus
KBP RefKB
Freebase
Factorie NLP
segmentation,
tagging,
parsing,
mentions,
coreference
Cross-doc Coreference
Universal Schema
(Matrix completion)
String-Slot
Extraction
Aggregation & Filtering
Output KB

String-Slot Extraction

String-Slot Extraction

- Lexicons created from Freebase
  - Religion: /religion
  - Cause of Death: /people/cause_of_death
  - Charges: /base/fight/crime_type
  - Titles: /business/job_title

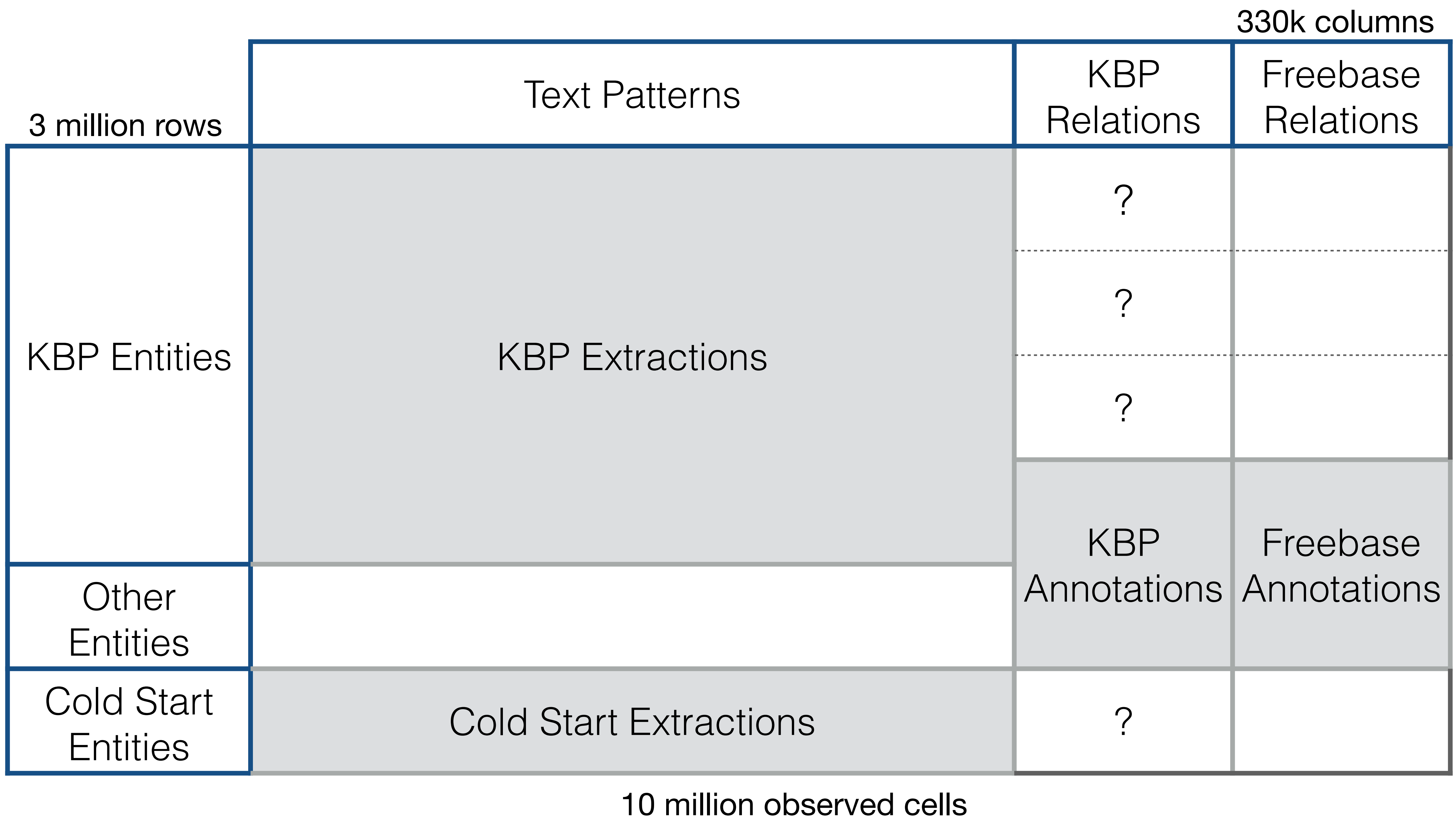
String-Slot Extraction

- Lexicons created from Freebase
  - Religion: /religion
  - Cause of Death: /people/cause_of_death
  - Charges: /base/fight/crime_type
  - Titles: /business/job_title
- Triggers for various slots
  - Date of birth/death: "*born*", "*died*", "*dies*"
  - Date founded: "*created*", "*founded*", "*established*"

String-Slot Extraction

- Lexicons created from Freebase
  - Religion: /religion
  - Cause of Death: /people/cause_of_death
  - Charges: /base/fight/crime_type
  - Titles: /business/job_title
- Triggers for various slots
  - Date of birth/death: "*born*", "*died*", "*dies*"
  - Date founded: "*created*", "*founded*", "*established*"
- Use Natty library to extract relative and absolute dates

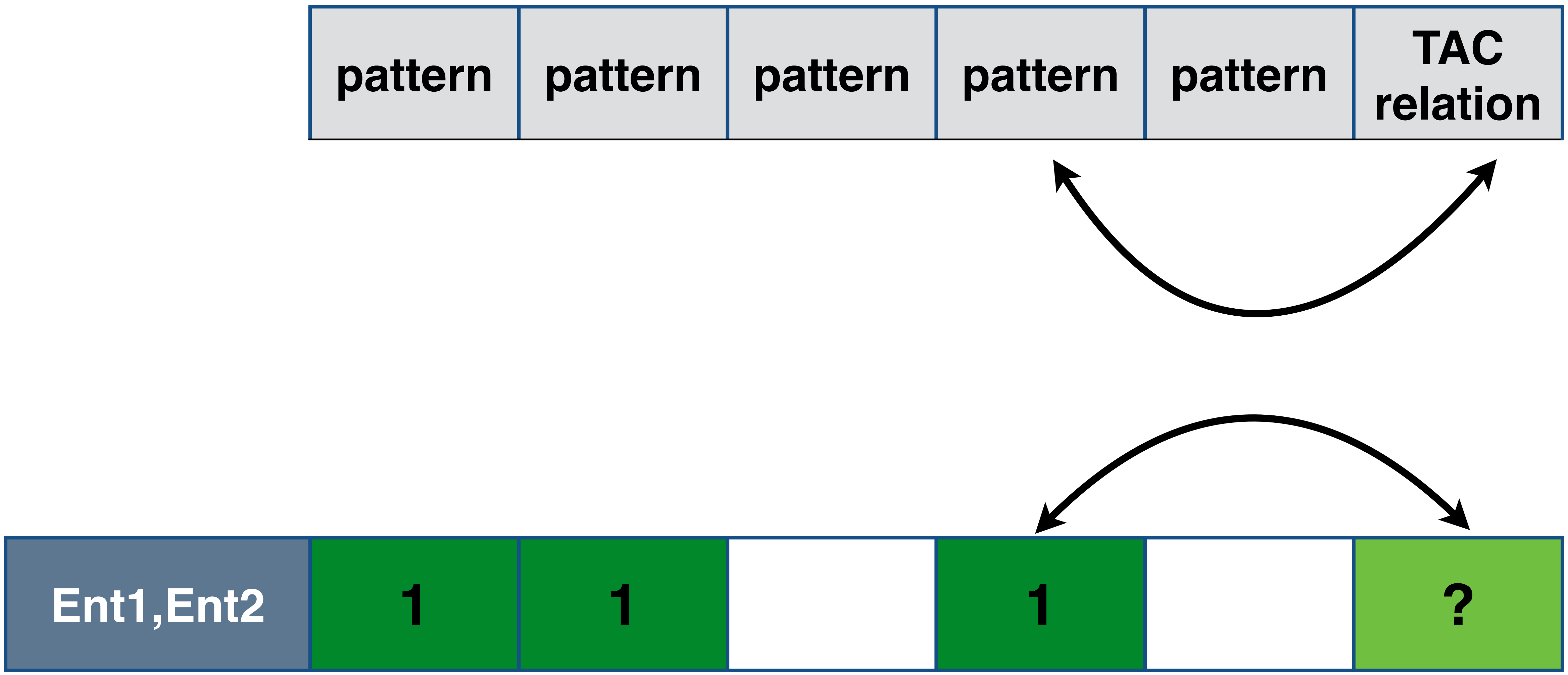
Matrix for KBP



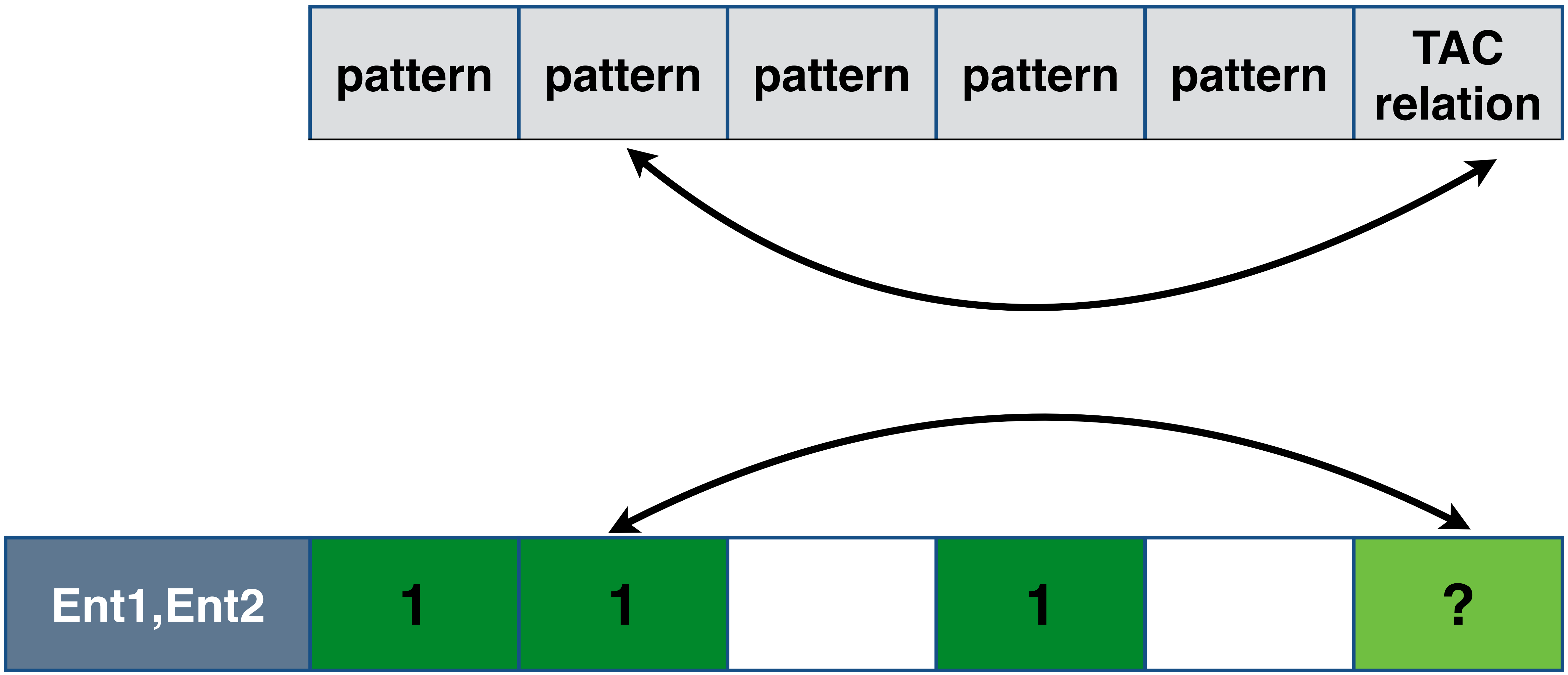
Provenance

	pattern	pattern	pattern	pattern	pattern	TAC relation
Ent1,Ent2	1	1		1		?

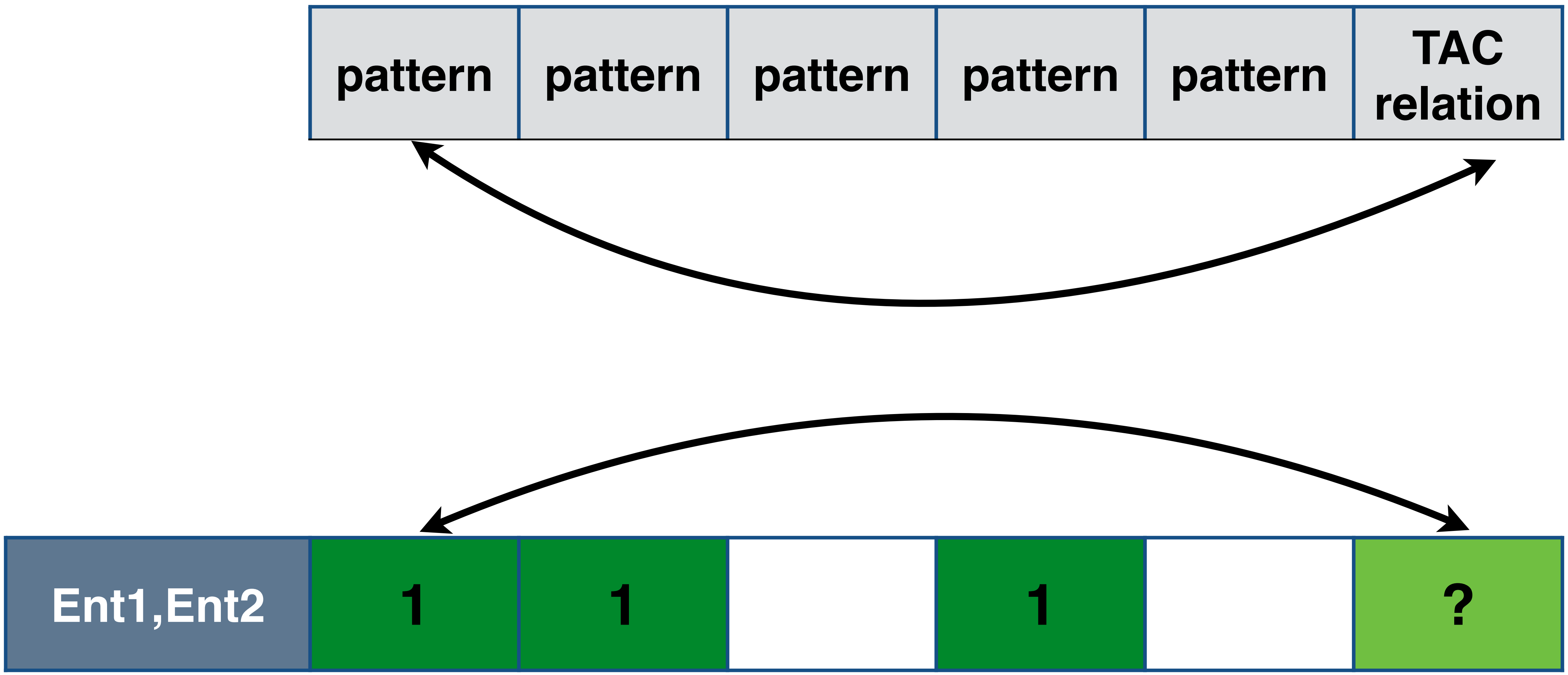
Provenance


pattern
pattern
pattern
pattern
pattern
TAC relation
Ent1,Ent2
1
1
1
?

Provenance

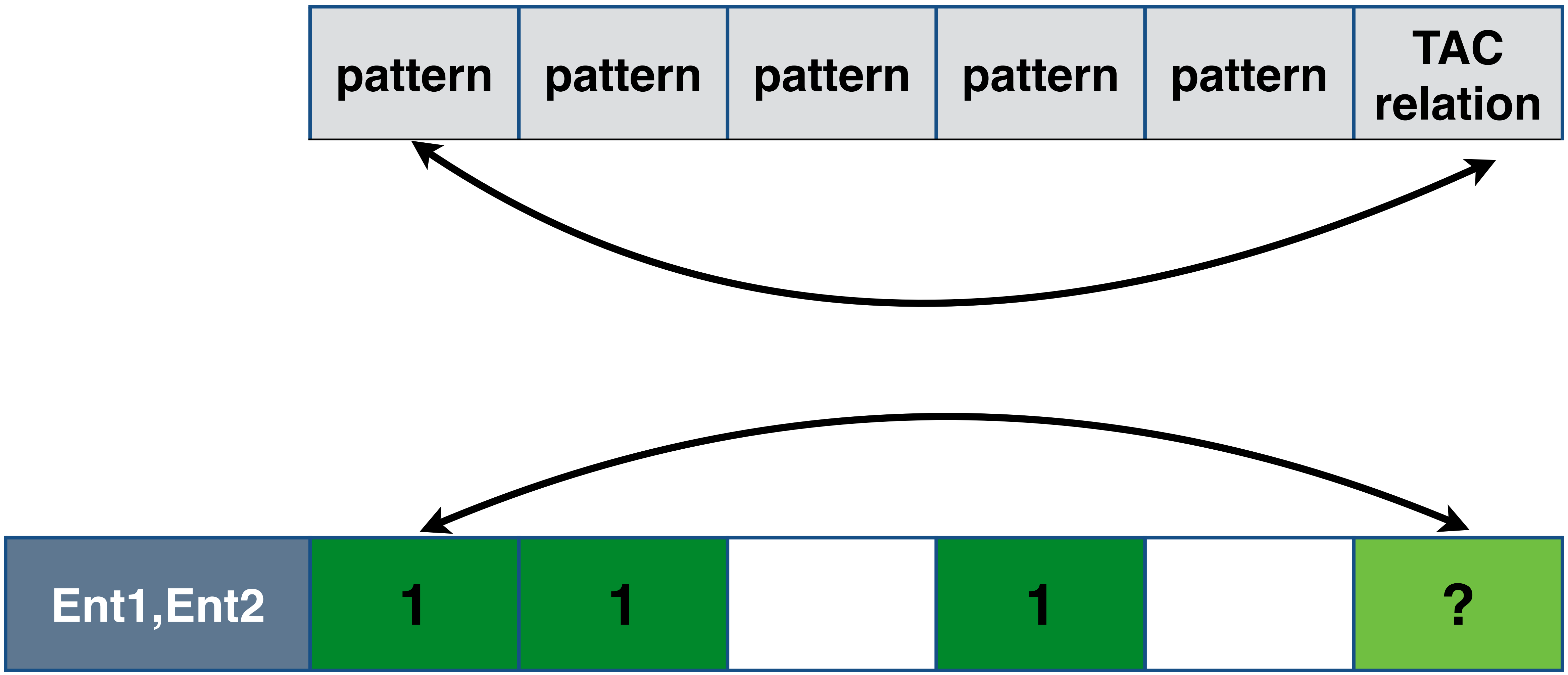

pattern
pattern
pattern
pattern
pattern
TAC relation
Ent1,Ent2
1
1
1
?

Provenance



pattern
pattern
pattern
pattern
pattern
TAC relation
Ent1,Ent2
1
1
1
?

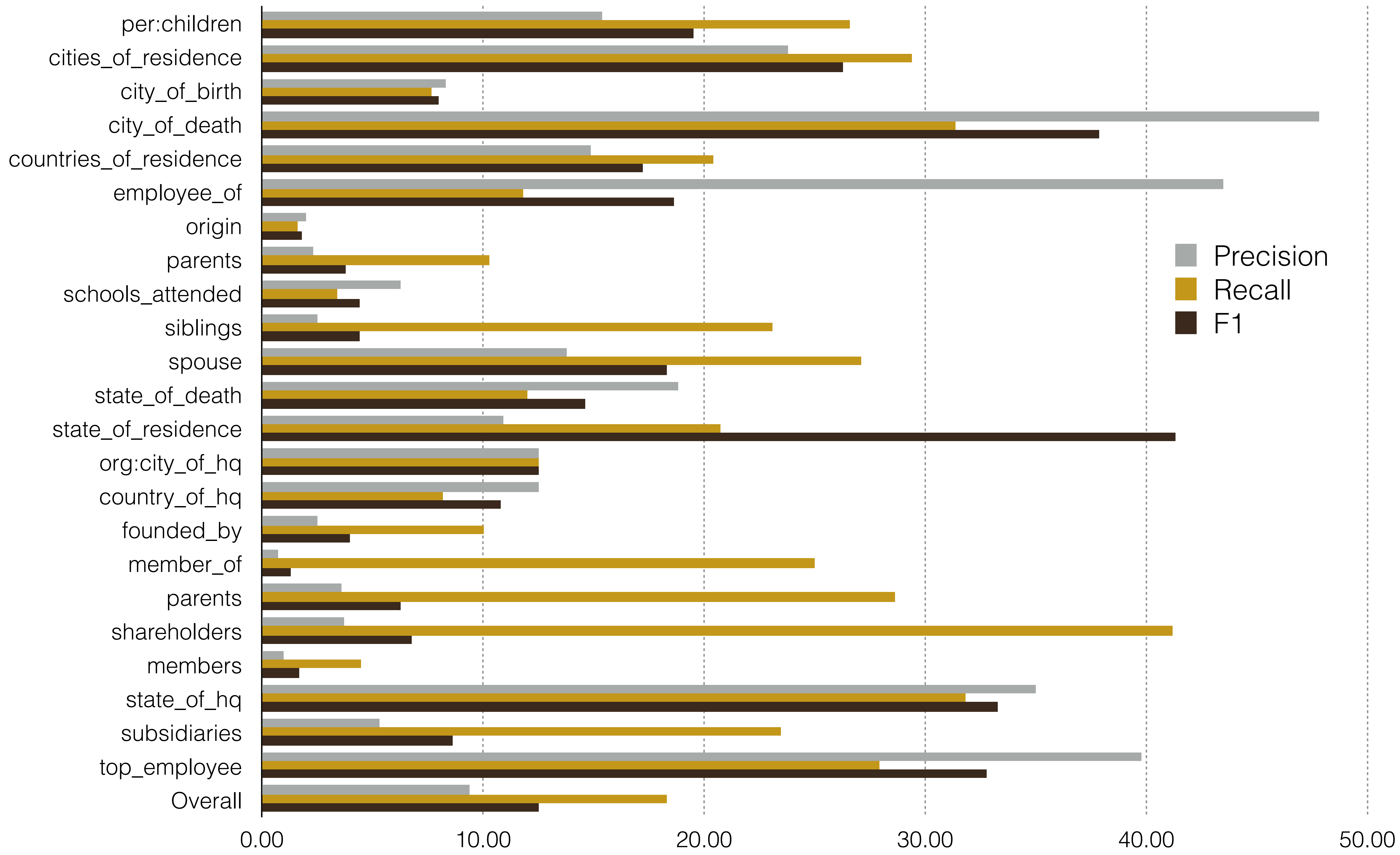
Provenance



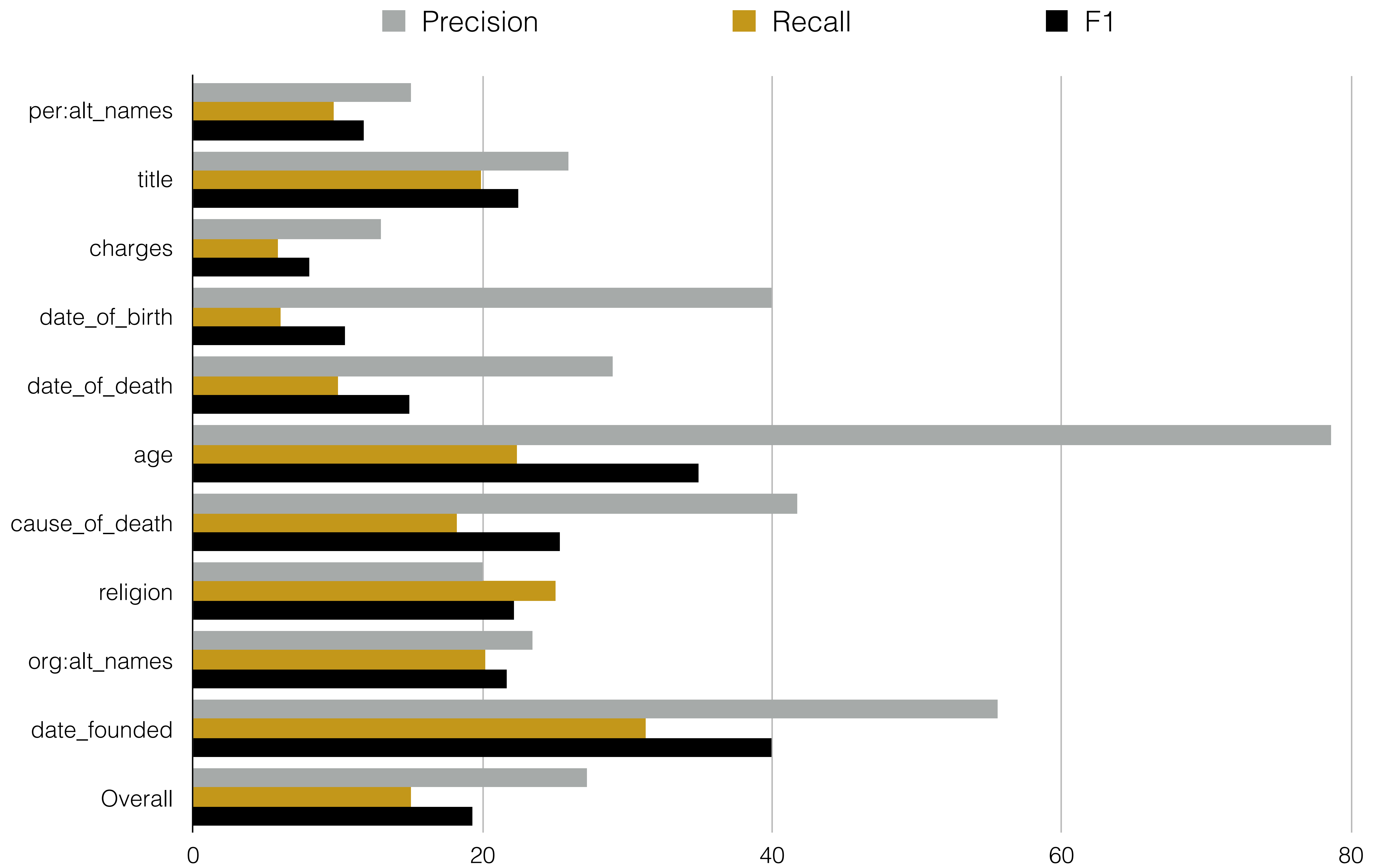
Pick Provenance

KBP 2013 Results

Entity-Based Slots



String-Based Slots



Overall Results

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1
0-hop	25.24	5.06	8.44

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1
0-hop	25.24	5.06	8.44
1-hop	11.42	3.46	5.32

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1
0-hop	25.24	5.06	8.44
1-hop	11.42	3.46	5.32
Combined	**17.88**	**4.38**	**7.03**

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1
0-hop	25.24	5.06	8.44
1-hop	11.42	3.46	5.32
Combined	**17.88**	**4.38**	**7.03**
NYU1	35.15	4.99	8.74

Overall Results

English Slot Filling Task

Type of Slots	Precision	Recall	F1
String-Slots	27.2	15.0	19.3
Entity-Slots	9.4	18.3	12.5
Overall (11th)	**18.5**	**10.9**	**13.7**
Top-10	30.3	9.67	14.6
Top-1	42.5	33.2	37.3

Cold Start Task

Evaluation	Precision	Recall	F1
0-hop	25.24	5.06	8.44
1-hop	11.42	3.46	5.32
Combined	**17.88**	**4.38**	**7.03**
NYU1	35.15	4.99	8.74
hltcoe4	34.3	23.39	27.81

Pattern Clusters

Pattern Clusters

cityOfResidence

- citizen of -
- on death row in -
- who be a native of -
- live in -

Pattern Clusters

cityOfResidence

- citizen of -
- on death row in -
- who be a native of -
- live in -

spouse

X's ex-husband, Y
- file for divorce from -
- hubby -
- accompany by his wife -
X, Y's ex-wife
- photographed with fiancee -
X, Y's girlfriend

Pattern Clusters

cityOfResidence

- citizen of -
- on death row in -
- who be a native of -
- live in -

employeeOf

- pay by state of -
- cartoonist at -
- replace NNP as face of -
- executive who run -
organization_founded
who resign from

spouse

X's ex-husband, Y
- file for divorce from -
- hubby -
- accompany by his wife -
X, Y's ex-wife
- photographed with fiancee -
X, Y's girlfriend

Pattern Clusters

cityOfResidence

- citizen of -
- on death row in -
- who be a native of -
- live in -

employeeOf

- pay by state of -
- cartoonist at -
- replace NNP as face of -
- executive who run -
organization_founded
who resign from

schoolsAttended

- from his/her days in college at -
- hockey player for -
- be graduate of -
- be junior at -
- have resign in protest from -
- lawyer be hire by -
employee_or_member_of
who resign from

spouse

X's ex-husband, Y
- file for divorce from -
- hubby -
- accompany by his wife -
X, Y's ex-wife
- photographed with fiancee -
X, Y's girlfriend

Conclusions

Summary

Summary

- Constructing a **Unified KB** is beneficial

Summary

- Constructing a **Unified KB** is beneficial
  - correlation-based propagation of supervision

Summary

- Constructing a **Unified KB** is beneficial
  - correlation-based propagation of supervision
  - learn implications between patterns & relations

Summary

- Constructing a **Unified KB** is beneficial
    - correlation-based propagation of supervision
    - learn implications between patterns & relations
- **Universal Schema** via Matrix Factorization

Summary

- Constructing a **Unified KB** is beneficial
  - correlation-based propagation of supervision
  - learn implications between patterns & relations
- **Universal Schema** via Matrix Factorization
  - provides an accurate embedding model

Summary

- Constructing a **Unified KB** is beneficial
  - correlation-based propagation of supervision
  - learn implications between patterns & relations
- **Universal Schema** via Matrix Factorization
  - provides an accurate embedding model
  - stochastic gradient descent training is scalable

Summary

- Constructing a **Unified KB** is beneficial
  - correlation-based propagation of supervision
  - learn implications between patterns & relations
- **Universal Schema** via Matrix Factorization
  - provides an accurate embedding model
  - stochastic gradient descent training is scalable
  - can be used for predicting new relations

Future Work

Future Work

- Use and predict **entity types** jointly

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model
- **Extend the model** to use:

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model
- **Extend the model** to use:
  - more freebase facts, features,

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model
- **Extend the model** to use:
  - more freebase facts, features,
  - all of the KBP source corpus

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model
- **Extend the model** to use:
  - more freebase facts, features,
  - all of the KBP source corpus
- **Better entity linking** for both within- and cross-document

Future Work

- Use and predict **entity types** jointly
  - use latent representation of entities for types
  - incorporate additional resources such as NELL, lexicons
- Incorporate **prior-knowledge** into the model
- **Extend the model** to use:
  - more freebase facts, features,
  - all of the KBP source corpus
- **Better entity linking** for both within- and cross-document
- Participate in **TAC KBP 2014**!

Thank you!