Cold Start KB and Slot-Filling Approaches

UMass Amherst

Ben Roth, Nick Monath, David Belanger,
Emma Strubell, Pat Verga and Andrew McCallum

Outline

- **Prediction Modules**
    - **Universal Schema**
    - **CNNs**
    - SVMs
    - Rule-based
- Slot-Filling vs. KB architectures
    - Entity expansion
    - Entity linking
- Multi-hop queries and Precision

Universal Schema



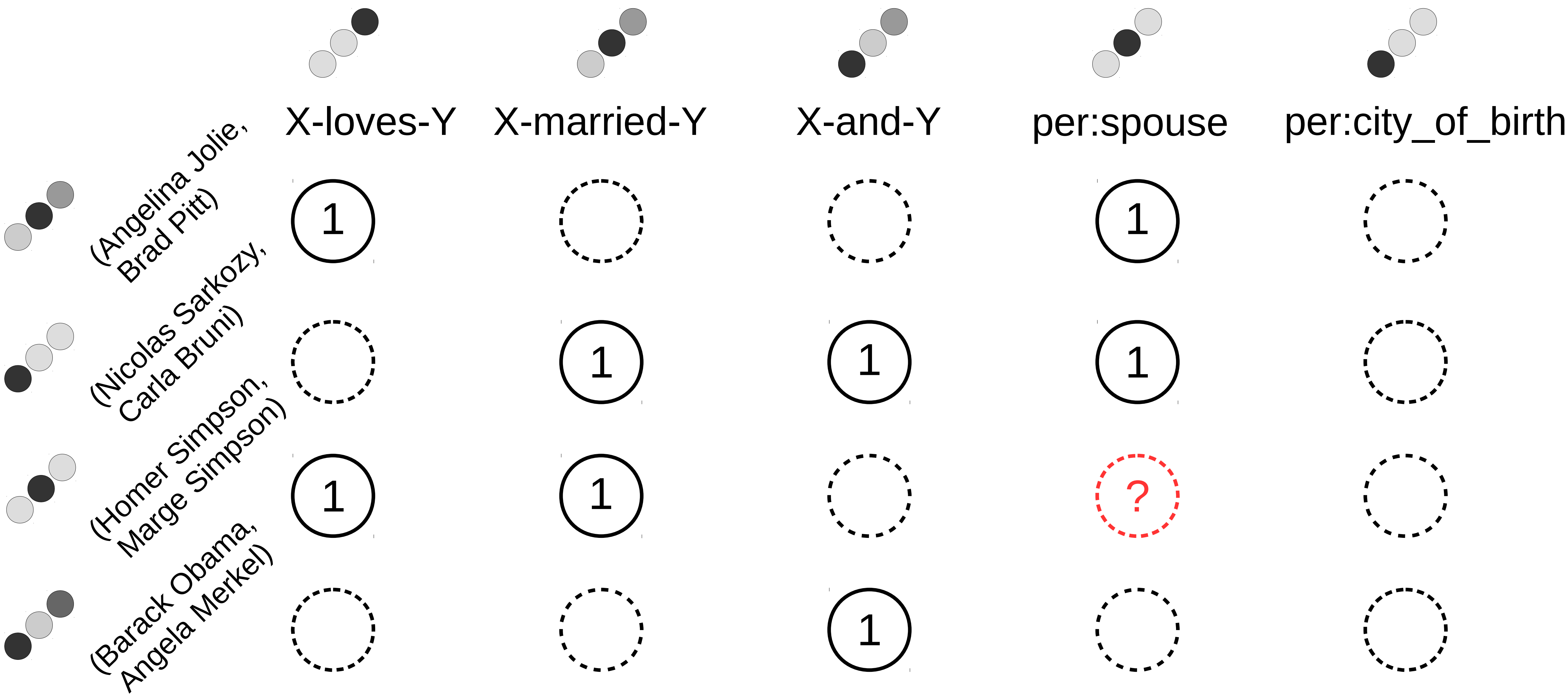

	X-loves-Y	X-married-Y	X-and-Y	per:spouse	per:city_of_birth
(Angelina Jolie, Brad Pitt)	1			1	
(Nicolas Sarkozy, Carla Bruni)		1	1	1	
(Homer Simpson, Marge Simpson)	1	1			
(Barack Obama, Angela Merkel)			1		

[Riedel et al., 2013]

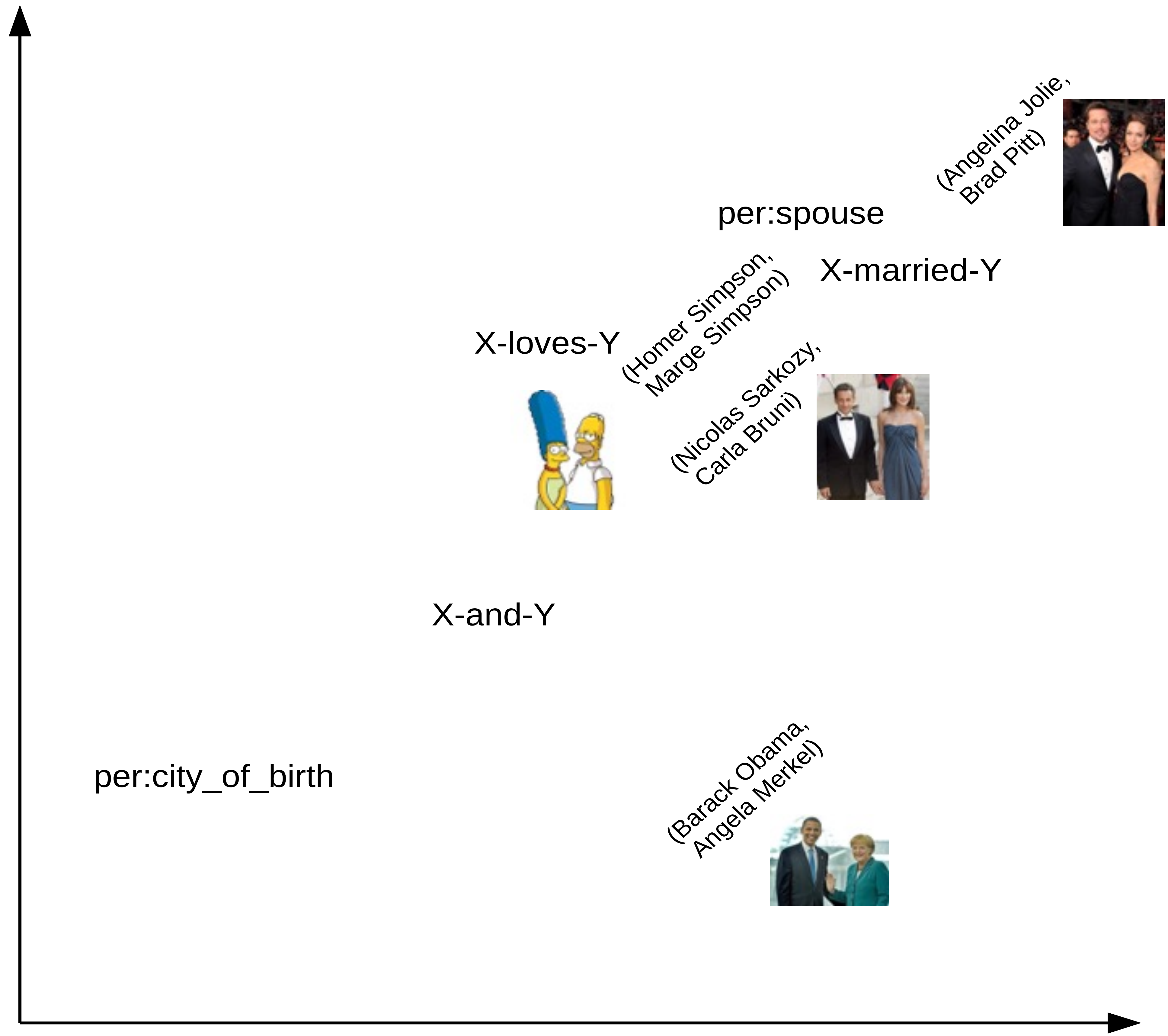
Universal Schema

	X-loves-Y	X-married-Y	X-and-Y	per:spouse	per:city_of_birth
(Angelina Jolie, Brad Pitt)	1			1	
(Nicolas Sarkozy, Carla Bruni)		1	1	1	
(Homer Simpson, Marge Simpson)	1	1			
(Barack Obama, Angela Merkel)			1		

Universal Schema



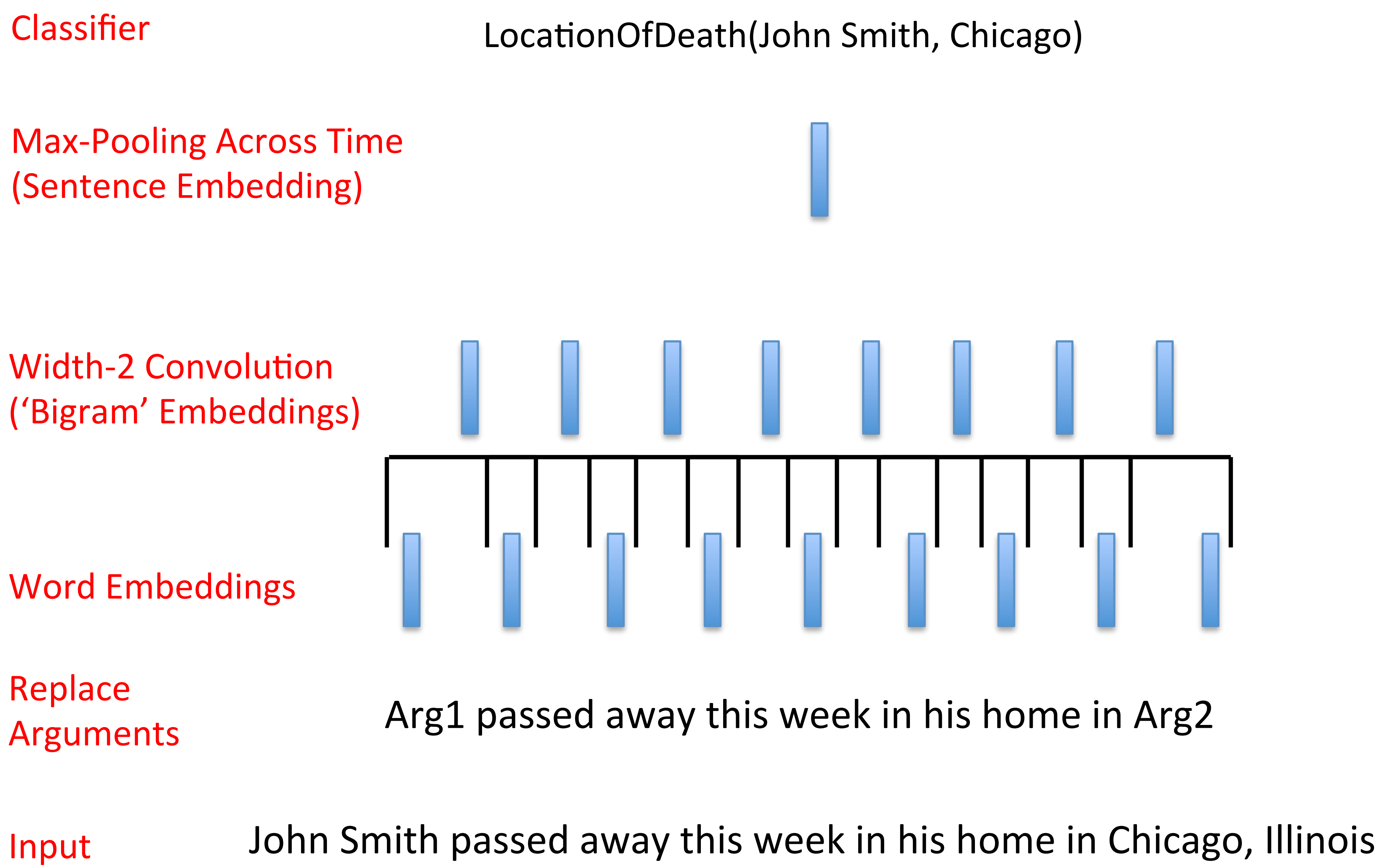
Universal Schema



Universal Schema & Convolutional Neural Nets

- Universal Schema
  - (+) Induces **smooth similarity measure** between context patterns and relations
  - (+) makes use of co-occurrences of the whole corpus (Even if no direct distant supervision match)
  - (-) **Entity pairs** only represented as **aggregates**, not mentions
  - (-) Contexts are atomic units
    [PER] passed away in [LOC]
- Convolutional Neural Network
  - related work:
    [Collobert et al., 2011], [Kalchbrenner et al, 2014], [Zeng et al., 2014, 2015], [Zhang and Wallace, 2015]
  - (+) Allow for **fine-grained analysis** of mention contexts
    - 'soft ngram' features
      [PER] passed away this week in his home **in [LOC]**
    - ngram features are known to perform well on KBP
  - (-) Requires sentence level distant supervision alignment

Relation Prediction with Convolutional Neural Nets



Outline

- **Prediction Modules**
  - Universal Schema
  - CNNs
  - **SVMs**
  - **Rule-based**
- Slot-Filling vs. KB architectures
  - Entity expansion
  - Entity linking
- Multi-hop queries and Precision

Support Vector Machines and Rule Based Modules

- **SVM Module**
  - Set of Binary Support Vector Machine Classifiers
  - Sparse n-gram features
  - Trained on distant supervision data
- **Hand-written Rules Module**
  - [ARG1] was born in [ARG2]
- **Alternate Names Module**
  - Rules based on Wikipedia anchor text statistics

Single Modules Comparison

	Prec	Rec	F1
USchema	26.54	**8.93**	**13.37**
SVM	27.09	8.80	13.29
CNN	16.45	5.54	8.29
Rules	**76.32**	3.75	7.16
all	14.68	13.44	14.03
w/o CNN	**22.32**	14.43	**17.53**
all*ignoretags	9.01	**16.5**	11.65

Ablation Analysis

	Prec	Rec	F1
all	14.68	13.44	14.03
w/o CNN	22.32	14.43	17.53
w/o USchema	11.5	12.91	12.16
w/o SVM	17.16	11.89	14.05
w/o Rules	10.76	11.94	11.32

Outline

- Prediction Modules
  - Universal Schema
  - CNNs
  - SVMs
  - Rule-based
- **Slot-Filling vs. KB architectures**
  - Entity expansion
  - Entity linking
- Multi-hop queries and Precision

Slot-Filling vs. KB Pipeline

- **Same prediction modules** for both settings
- Only difference is in **query expansion** and **entity linking**
- Slot Filling:
  - **Iterative query-based retrieval**
  - Query is expanded and matched in documents
- KB Construction:
  - **Knowledge-base is constructed ahead of time**
  - All entities found by the NE-Tagger are linked or clustered

Slot-Filling Pipeline

Query

Name: “Facebook”
Slot0: org:subsidiaries
Slot1: org:founders

“Facebook, Inc.” “facebook.com”

Slot-Filling Pipeline

Query

Name: "Facebook"
Slot0: org:subsidiaries
Slot1: org:founders

"Facebook, Inc." "facebook.com"

... reminiscent of **Instagram**'s parent company **Facebook Inc.** ...
... the $19 billion buyout of **Whatsapp** by **Facebook** ...

Slot-Filling Pipeline

Query

Name: "Facebook"
Slot0: org:subsidiaries
Slot1: org:founders

"Facebook, Inc." "facebook.com"

... reminiscent of **Instagram**'s parent company **Facebook Inc.** ...
... the $19 billion buyout of **Whatsapp** by **Facebook** ...

ARG1	rel	ARG2
Facebook	org:subsidiaries	Instagram
Facebook	org:subsidiaries	Whatsapp

Slot-Filling Pipeline

Query

Name: “Facebook”
Slot0: org:subsidiaries
Slot1: org:founders

“Instagram”

ARG1	rel	ARG2
Facebook	org:subsidiaries	Instagram
Facebook	org:subsidiaries	Whatsapp

Slot-Filling Pipeline

Query

Name: “Facebook”
Slot0: org:subsidiaries
Slot1: org:founders

“Instagram” 

... prior to founding **Instagram**, **Kevin Systrom** was of the startup ...
... **Mike Krieger** co-founded **Instagram** with Kevin Systrom ...

ARG1	rel	ARG2
Facebook	org:subsidiaries	Instagram
Facebook	org:subsidiaries	Whatsapp

Slot-Filling Pipeline

Query

Name: “Facebook”
Slot0: org:subsidiaries
Slot1: org:founders

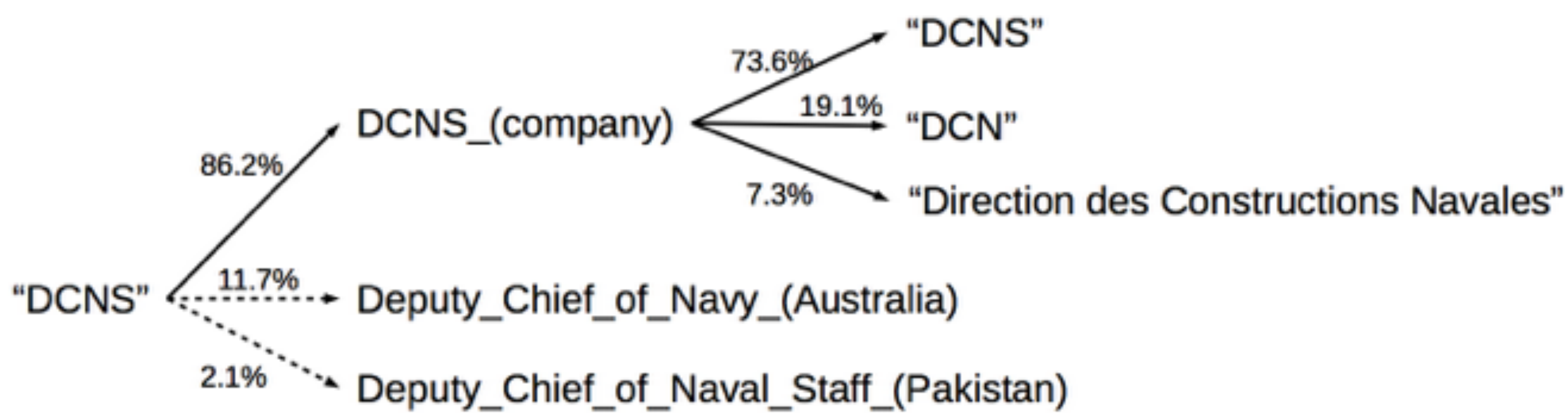
“Instagram”

... prior to founding **Instagram**, **Kevin Systrom** was of the startup ...
... **Mike Krieger** co-founded **Instagram** with Kevin Systrom ...

ARG1	rel	ARG2
Facebook	org:subsidiaries	Instagram
Facebook	org:subsidiaries	Whatsapp
Instagram	org:founders	Kevin Systrom
Instagram	org:founders	Mike Krieger

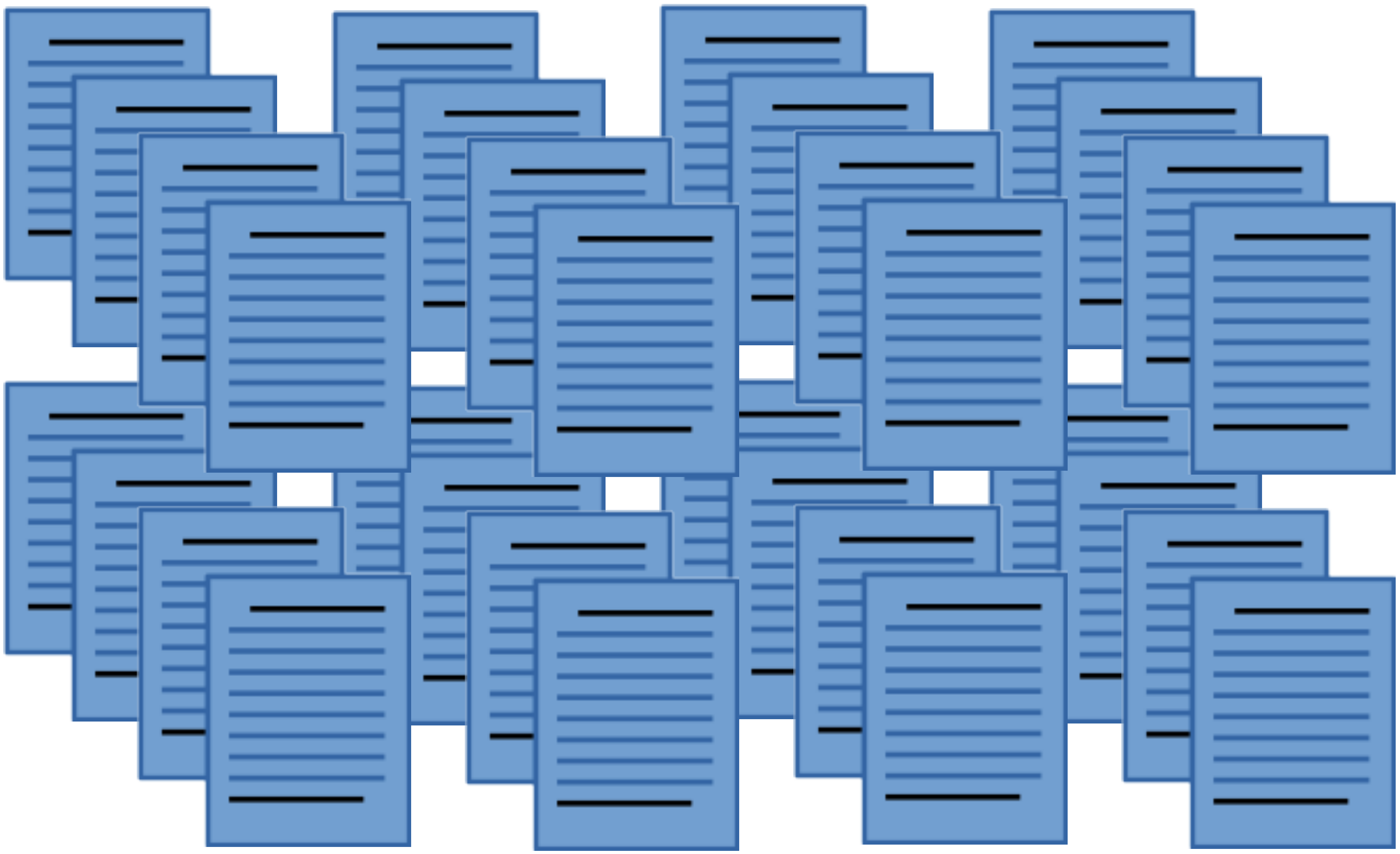
SF Setting: Entity Expansion

- Retrieval pipeline **controls precision and recall**
- **Expand query** to most likely anchor texts (**recall**)

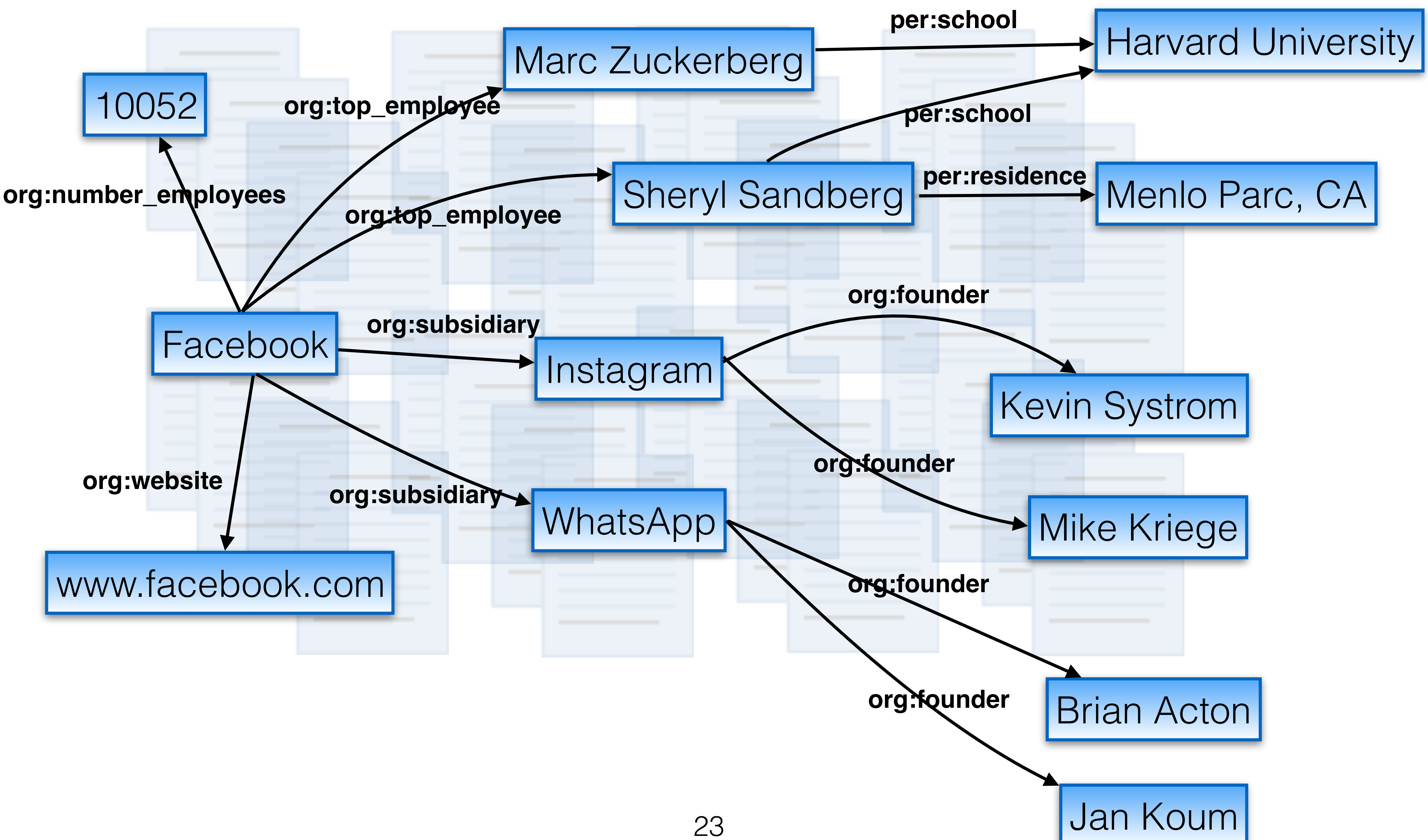


- Find **single best expansion** for document retrieval (**precision**)
  - PPMI on document collection
- After retrieval, use **all expansions** for query matching (**recall**)

KB Pipeline

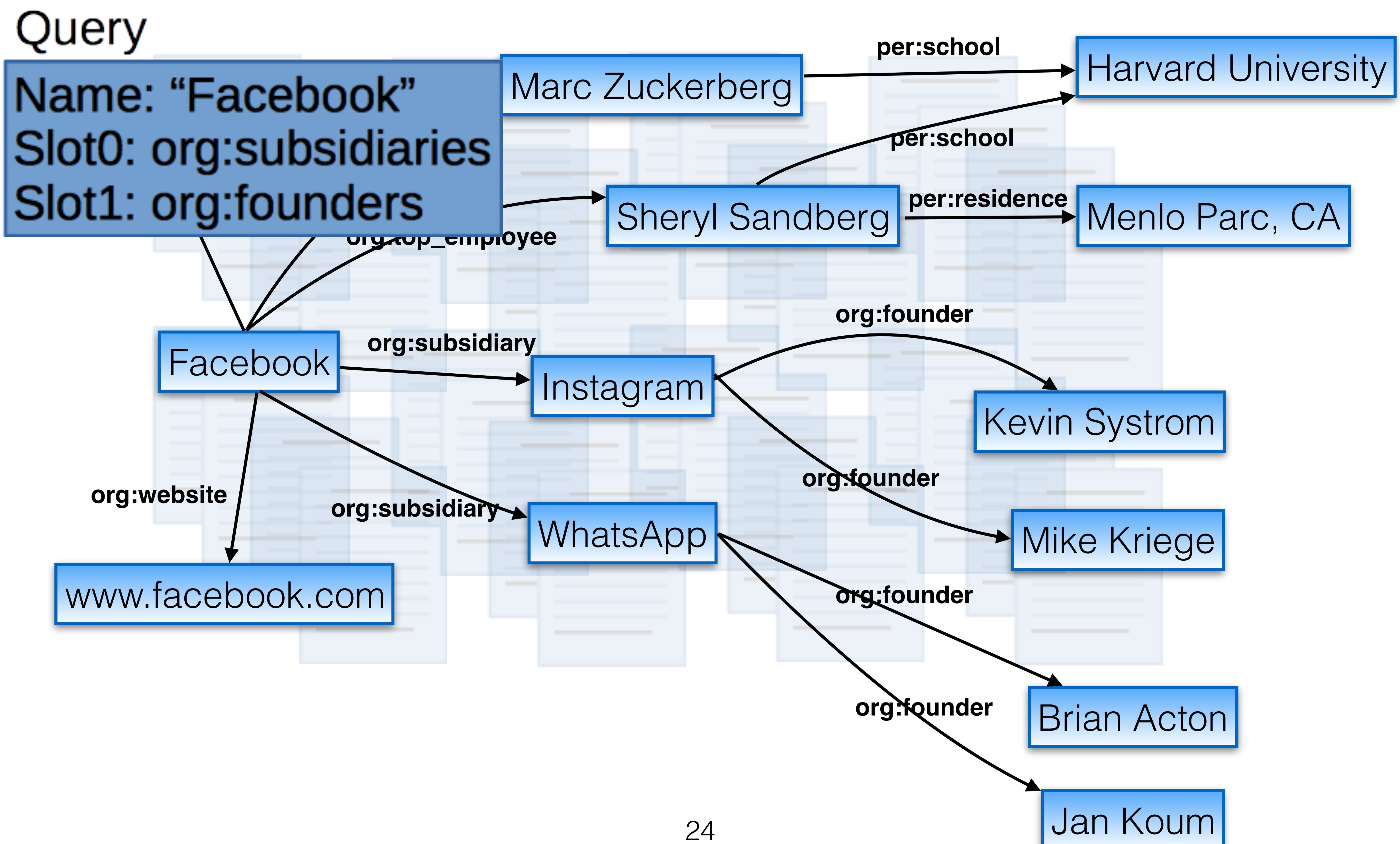


KB Pipeline



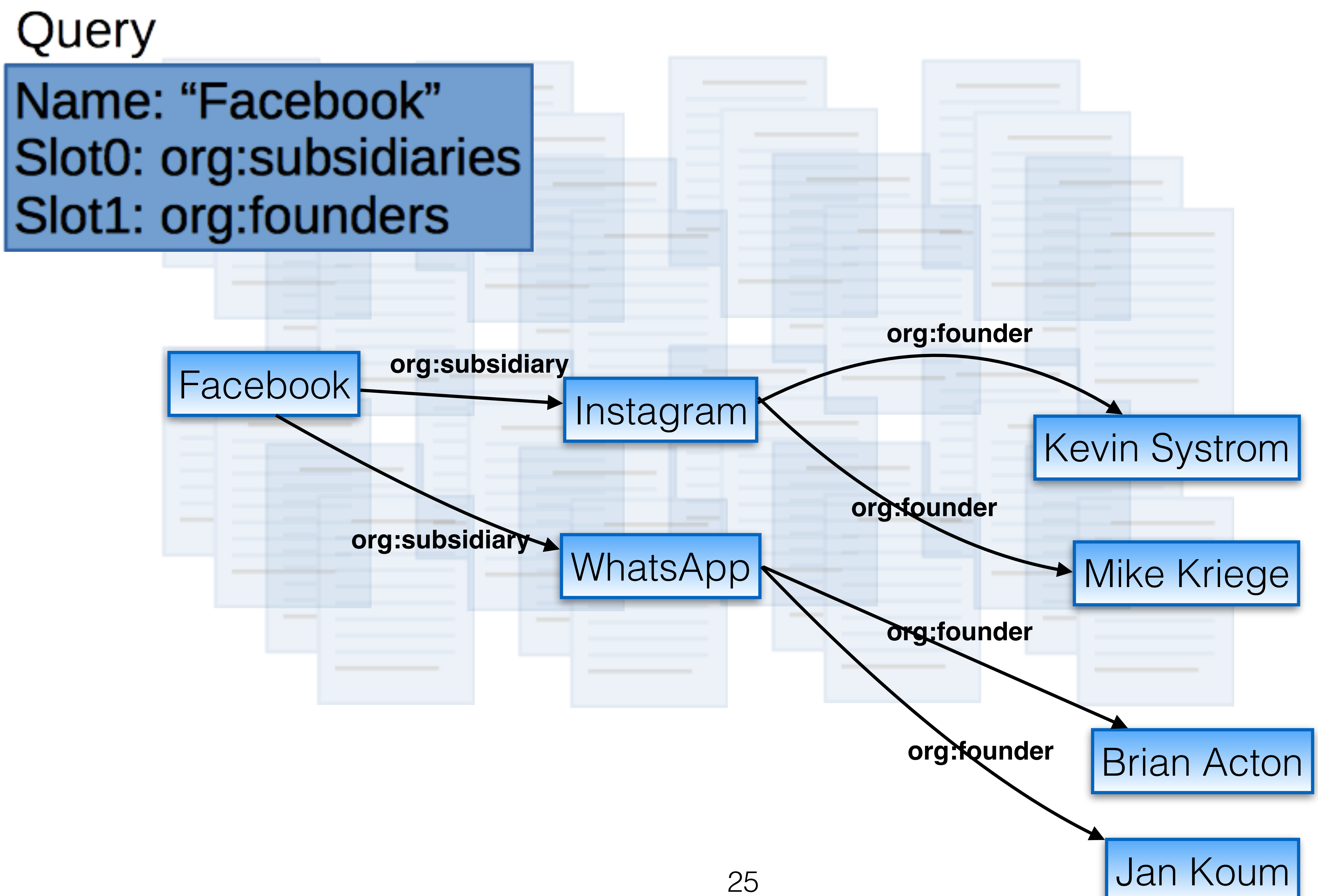
Marc Zuckerberg
per:school
Harvard University
10052
org:top_employee
per:school
org:number_employees
Sheryl Sandberg
per:residence
Menlo Parc, CA
org:top_employee
org:founder
Facebook
org:subsidiary
Instagram
Kevin Systrom
org:founder
org:website
org:subsidiary
WhatsApp
Mike Kriege
www.facebook.com
org:founder
org:founder
Brian Acton
Jan Koum

KB Pipeline



Query
Name: “Facebook”
Slot0: org:subsidiaries
Slot1: org:founders
Marc Zuckerberg
per:school
Harvard University
per:school
Sheryl Sandberg
per:residence
Menlo Parc, CA
Facebook
org:subsidiary
Instagram
org:founder
Kevin Systrom
org:founder
Mike Kriege
org:website
org:subsidiary
WhatsApp
www.facebook.com
org:founder
Brian Acton
org:founder
Jan Koum

KB Pipeline


Query
Name: "Facebook"
Slot0: org:subsidiaries
Slot1: org:founders
Facebook
org:subsidiary
Instagram
org:subsidiary
WhatsApp
org:founder
Kevin Systrom
org:founder
Mike Kriege
org:founder
Brian Acton
org:founder
Jan Koum

KB Setting: Entity Linking

The American Federation of Teachers and the Boston Teachers Union, its local affiliate, have now demonstrated why they should be viewed through those skeptical spectacles. The BTU leadership urged its members to back Marty Walsh. The American Federation of Teachers, the BTU 's parent, was clandestinely scheming to elect Walsh and defeat John Connolly, a pointed BTU critic. Walsh shouldn't be blamed for the AFT 's electoral subterfuge. During his campaign, Walsh portrayed himself as intent on bringing change to the Boston schools.

KB Setting: Entity Linking

- Perform within-doc coref & select canonical mention
- retrieve Wikipedia articles based on anchor text

The **American Federation of Teachers** and the **Boston Teachers Union**, its local affiliate, have now demonstrated why they should be viewed through those skeptical spectacles. The **BTU** leadership urged its members to back **Marty Walsh**. The **American Federation of Teachers**, the **BTU** 's parent, was clandestinely scheming to elect **Walsh** and defeat **John Connolly**, a pointed **BTU** critic. **Walsh** shouldn't be blamed for the **AFT** 's electoral subterfuge. During his campaign, **Walsh** portrayed himself as intent on bringing change to the **Boston** schools.

KB Setting: Entity Linking

- Perform within-doc coref & select canonical mention
- retrieve Wikipedia articles based on anchor text

The **American Federation of Teachers** and the **Boston Teachers Union**, its local affiliate, have now demonstrated why they should be viewed through those skeptical spectacles. The **BTU** leadership urged its members to back **Marty Walsh**. The **American Federation of Teachers**, the **BTU** 's parent, was clandestinely scheming to elect **Walsh** and defeat **John Connolly**, a pointed **BTU** critic. **Walsh** shouldn't be blamed for the **AFT** 's electoral subterfuge. During his campaign, **Walsh** portrayed himself as intent on bringing change to the **Boston** schools.

Context Vector

WIKIPEDIA

KB Setting: Entity Linking

- compute cosine similarity to current TAC document
- if threshold is exceeded link to article with highest similarity

The **American Federation of Teachers** and the **Boston Teachers Union**, its local affiliate, have now demonstrated why they should be viewed through those skeptical spectacles. The **BTU** leadership urged its members to back **Marty Walsh**. The **American Federation of Teachers**, the **BTU** 's parent, was clandestinely scheming to elect **Walsh** and defeat **John Connolly**, a pointed **BTU** critic. **Walsh** shouldn't be blamed for the **AFT** 's electoral subterfuge. During his campaign, **Walsh** portrayed himself as intent on bringing change to the **Boston** schools.

Context Vector

WIKIPEDIA

SF vs KB Pipeline Results

	Prec	Rec	F1
UMass_SF	**20.20**	13.20	**15.97**
UMass_KB	10.33	**14.17**	11.95

- SF and KB use same prediction modules (USchema+SVM)
- Only difference is linking/expansion
- => results underline importance of entity linking

Outline

- Prediction Modules
    - Universal Schema
    - CNNs
    - SVMs
    - Rule-based
- Slot-Filling vs. KB architectures
    - Entity expansion
    - Entity linking
- **Multi-hop queries and Precision**

Precision, Multi-hop queries…

	1-hop queries Prec	2-hop queries Prec	1-hop queries F1	2-hop queries F1
UMass SF1	33.27%	8.91%	23.51%	8.11%
UMass SF5	31.75%	7.64%	21.79%	7.24%
UMass KB1 (SF5 equiv)	22.66%	3.76%	19.15%	5.41%

Precision, Multi-hop queries…
… and the Right to Remain Silent

	submission			not predicting 2-hop queries		
Run	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**
SF1	**0.2232**	0.1443	**0.1753**	**0.3327**	0.1185	0.1747
SF2	0.0901	0.1650	0.1165	0.2175	**0.1321**	*0.1644*
SF3	0.2034	0.1528	0.1745	0.3172	0.1275	***0.1819***
SF4	0.2186	0.1159	0.1514	0.3200	0.0984	0.1505
SF5	0.2020	0.1320	0.1597	0.3175	0.1081	*0.1613*
KB1	0.1033	0.1417	0.1195	0.2266	0.0971	*0.1359*
KB2	0.0768	**0.1657**	0.1050	0.1729	0.1198	*0.1415*
KB3	0.0883	0.1139	0.0995	0.1895	0.0842	*0.1166*
KB4	0.1015	0.1204	0.1102	0.2070	0.0919	*0.1273*

Precision, Multi-hops…
… Man vs. Machine

	Prec 1-hop queries	Prec 2-hop queries	exponent $Prec_1^x=Prec_2$
Humans	85.38%	75.97%	**1.74**
UMass_SF1	33.27%	8.91%	2.19
Top1 system	50.15%	21.21%	2.24

Precision, Multi-hops

- Precision decays quadratically in the number of hops.

- Humans are (over-proportionally) better at jointly predicting chains of relations.

- ***Not*** predicting the second hop gives better results in 7 out of 9 settings!

- => results motivate research on KB reasoning approaches.

Conclusion

- Universal Schema and SVM strongest components
- Entity linking most important problem for KB setting
- Precision is lost on multi-hop queries
  - Better not to predict hop 2 at all …
  - Humans answer multi-hop queries jointly
  - Strong motivation for joint reasoning approaches