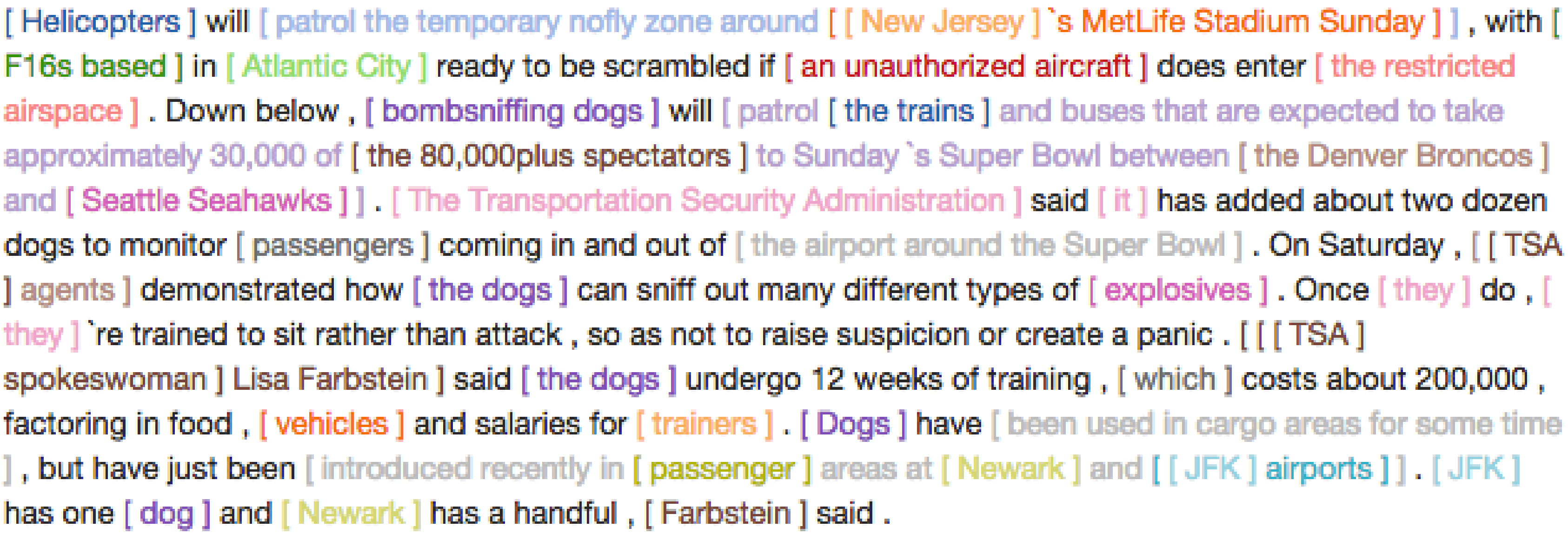
Coreference Resolution with Knowledge

Haoruo Peng

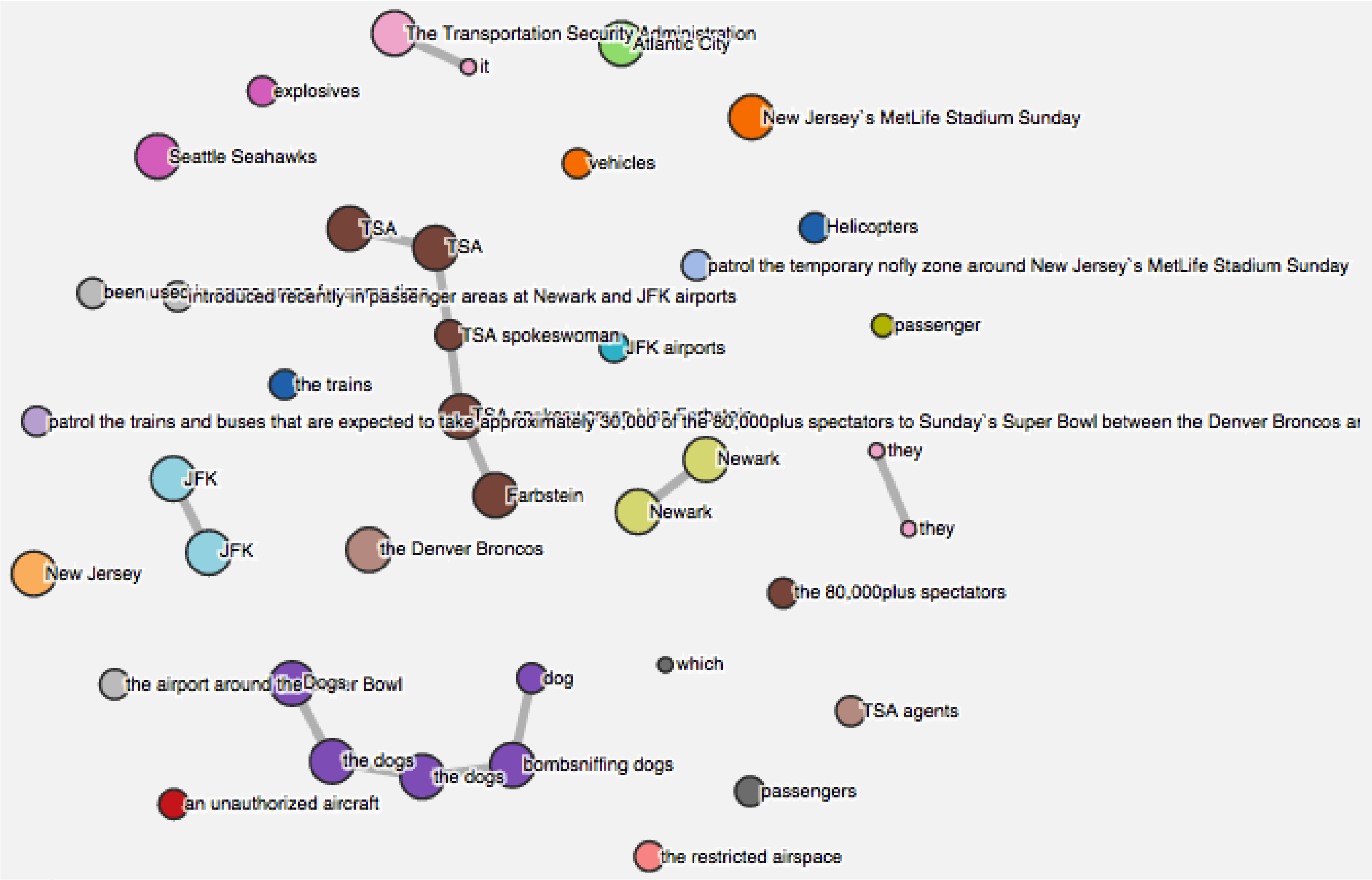
March 20, 2015

Coreference Problem



[Helicopters] will [patrol the temporary nofly zone around [[New Jersey] `s MetLife Stadium Sunday]] , with [F16s based] in [Atlantic City] ready to be scrambled if [an unauthorized aircraft] does enter [the restricted airspace] . Down below , [bombsniffing dogs] will [patrol [the trains] and buses that are expected to take approximately 30,000 of [the 80,000plus spectators] to Sunday `s Super Bowl between [the Denver Broncos] and [Seattle Seahawks]] . [The Transportation Security Administration] said [it] has added about two dozen dogs to monitor [passengers] coming in and out of [the airport around the Super Bowl] . On Saturday , [[TSA] agents] demonstrated how [the dogs] can sniff out many different types of [explosives] . Once [they] do , [they] `re trained to sit rather than attack , so as not to raise suspicion or create a panic . [[[TSA] spokeswoman] Lisa Farbstein] said [the dogs] undergo 12 weeks of training , [which] costs about 200,000 , factoring in food , [vehicles] and salaries for [trainers] . [Dogs] have [been used in cargo areas for some time] , but have just been [introduced recently in [passenger] areas at [Newark] and [[JFK] airports]] . [JFK] has one [dog] and [Newark] has a handful , [Farbstein] said .

Coreference Problem



Online Demo

- Please check outhttp://cogcomp.cs.illinois.edu/page/demo_view/Coref

General Framework

Jack threw the bags of Mary into the water since he is angry with her.

General Framework

Jack threw the bags of Mary into the water since he is angry with her.

Mention Detection

General Framework

Jack threw the bags of ***Mary*** into the water since ***he*** is angry with ***her***.

Mention Detection

General Framework

Jack threw the bags of ***Mary*** into the water since ***he*** is angry with ***her***.

Mention Detection

Pairwise Mention Scoring

General Framework

Jack threw the bags of ***Mary*** into the water since ***he*** is angry with ***her***.

- Mention Detection
- Pairwise Mention Scoring
  - Goal: Coreferent Mentions have higher scores

General Framework

Jack threw the bags of ***Mary*** into the water since ***he*** is angry with ***her***.

- Mention Detection
- Pairwise Mention Scoring
  - Goal: Coreferent Mentions have higher scores

General Framework

Jack threw the bags of ***Mary*** into the water since ***he*** is angry with ***her***.

- Mention Detection
- Pairwise Mention Scoring
  - Goal: Coreferent Mentions have higher scores
- Inference (ILP)
  - Best-Link
  - All-Link

ILP formulation of CR

- Best-Link

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv}$$

$$\text{s.t.} \quad \sum_{u<v} y_{uv} \leq 1, \ \forall v$$

$$y_{uv} \in \{0, 1\}$$

- All-Link

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv}$$

$$\text{s.t.} \quad y_{um} \geq y_{uv} + y_{vm} - 1, \ \forall u, v, m$$

$$y_{uv} \in \{0, 1\}$$

General Framework

- Learning (for Pairwise Mention Scores)
  - Structural SVM
  - Features:
    - Mention Types, String Relations, Semantic, Relative Location, Anaphoricity(Learned), Aligned Modifiers, Memorization, etc.

General Framework

- Learning (for Pairwise Mention Scores)
  - ☐ Structural SVM
  - ☐ Features:
    - Mention Types, String Relations, Semantic, Relative Location, Anaphoricity(Learned), Aligned Modifiers, Memorization, etc.

Evaluation

MUC

BCUB

CEAF_e

General Framework

- Learning (for Pairwise Mention Scores)
  - Structural SVM
  - Features:
    - Mention Types, String Relations, Semantic, Relative Location, Anaphoricity(Learned), Aligned Modifiers, Memorization, etc.

Evaluation

MUC

BCUB

CEAF_e

IllinoisCoref

VS. Stanford Muti-pass Sieve System
VS. Berkeley CR System

Difficulties in CR

- Hard Coreference Problems
  - [A bird] perched on the [limb] and [it] bent.
  - [Robert] is robbed by [Kevin], and [he] is arrested by police.

  - Gender / Plurality information cannot help
  - -> Requires Knowledge

Part 1
Solving Hard Coreference Problems

Hard Coreference Problems

- Motivating Examples

Hard Coreference Problems

- Motivating Examples
  - Category 1
    - [A bird] perched on the [limb] and [it] bent.
    - [The bee] landed on [the flower] because [it] had pollen.

Hard Coreference Problems

- Motivating Examples

  Category 1

  [A bird] perched on the [limb] and [it] bent.

  [The bee] landed on [the flower] because [it] had pollen.

  Category 2

  [Jack] is robbed by [Kevin], and [he] is arrested by police.

  [Jim] was afraid of [Robert] because [he] gets scared around new people.

Hard Coreference Problems

- Motivating Examples

Category 1

[A bird] perched on the [limb] and [it] bent.

[The bee] landed on [the flower] because [it] had pollen.

Category 2

[Jack] is robbed by [Kevin], and [he] is arrested by police.

[Jim] was afraid of [Robert] because [he] gets scared around new people.

C

[Lakshman] asked [Vivan] to get him some ice cream because [he] was hot.

Predicate Schemas

- Type 1
  -
- Type 2
  -

Predicate Schemas

- Type 1
  - $pred_m(m, a)$
- Type 2
  -

Predicate Schemas

- Type 1
  - $pred_m(\overset{\text{Sub}}{\boxed{m}}, \overset{\text{Obj}}{\boxed{a}})$
- Type 2

Predicate Schemas

- Type 1
  - $pred_m(\overset{\text{Sub}}{\boxed{m}}, \overset{\text{Obj}}{\boxed{a}})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.
- Type 2

Predicate Schemas

- Type 1
  - $pred_m(\overset{\text{Sub}}{\boxed{m}}, \overset{\text{Obj}}{\boxed{a}})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee], a=[pollen]))
- Type 2

Predicate Schemas

- Type 1
  - $pred_m(\overset{\text{Sub}}{\boxed{m}}, \overset{\text{Obj}}{\boxed{a}})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee], a=[pollen]))

- Type 2
  - $pred_m(m, a)|\hat{pred}_m(m, \hat{a}), cn$

Predicate Schemas

- Type 1
  - $pred_m(\overset{\text{Sub}}{\boxed{m}}, \overset{\text{Obj}}{\boxed{a}})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee], a=[pollen]))

- Type 2
  - $pred_m(m, a)|\widehat{pred}_m(m, \hat{a}), cn$

Predicate Schemas

Sub Obj

- Type 1
  - $pred_m(\boxed{m}, \boxed{a})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee], a=[pollen]))

Shared Mention

- Type 2
  - $pred_m(\boxed{m}, a) | \widehat{pred}_m(\boxed{m}, \hat{a}), cn$

Predicate Schemas

- Type 1

  Sub Obj

  - ☐ $pred_m(\boxed{m}, \boxed{a})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee], a=[pollen]))

  Shared Mention

- Type 2
  - ☐ $pred_m(\boxed{m}, a) | \widehat{pred}_m(\boxed{m}, \hat{a}), cn$

    (Cat2) [Jim] was afraid of [Robert] because [he] gets scared around new people.

Predicate Schemas

- Type 1
  - ☐ $pred_m(\underset{\text{Sub}}{\boxed{m}}, \underset{\text{Obj}}{\boxed{a}})$

    (Cat1) [The bee] landed on [the flower] because [it] had pollen.

    S(have(m=[the flower], a=[pollen])) >

    S(have(m=[the bee],   a=[pollen]))

- Type 2
  - ☐ $pred_m(\boxed{m}, a) | \widehat{pred}_m(\boxed{m}, \hat{a}), cn$ (Shared Mention)

    (Cat2) [Jim] was afraid of [Robert] because [he] gets scared around new people.

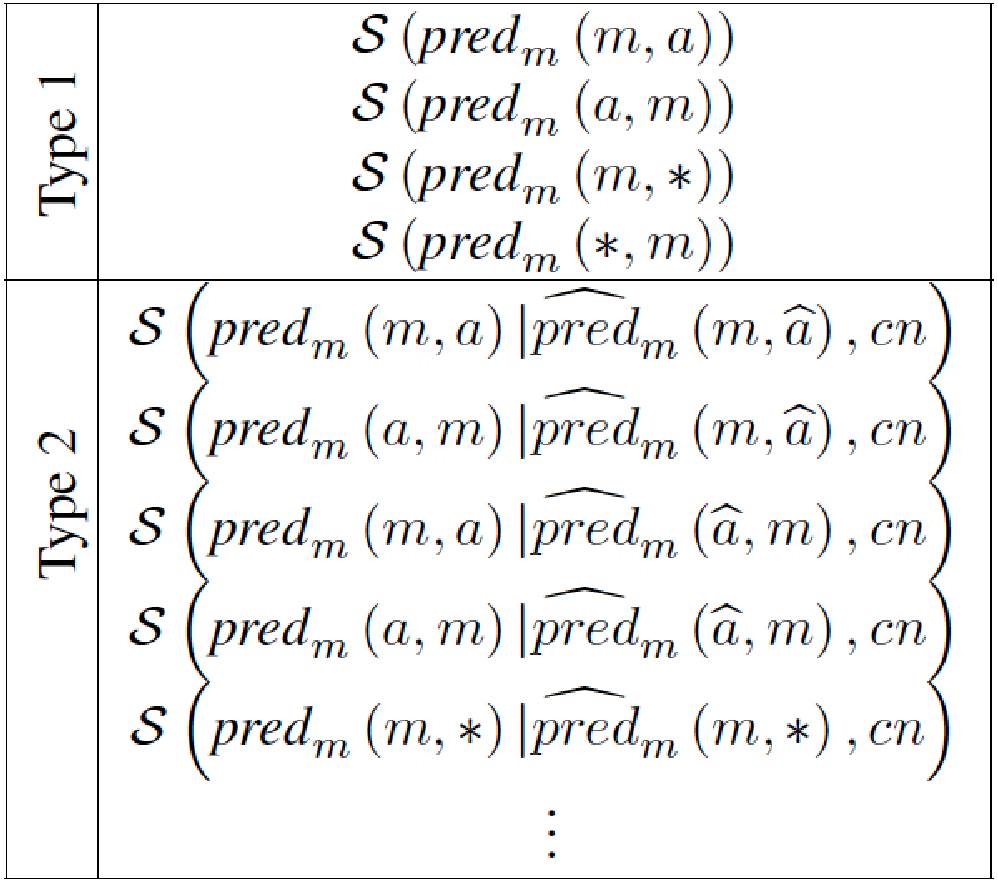
    ^

    S

    S(be afraid of(a=*, m=*) | get scared around(m=*, â=*), because)

Predicate Schemas

- Possible variations for scoring function statistics.



Type 1	$\mathcal{S}\left(pred_m\left(m,a\right)\right)$
	$\mathcal{S}\left(pred_m\left(a,m\right)\right)$
	$\mathcal{S}\left(pred_m\left(m,*\right)\right)$
	$\mathcal{S}\left(pred_m\left(*,m\right)\right)$
Type 2	$\mathcal{S}\left(pred_m\left(m,a\right) \mid \widehat{pred}_m\left(m,\hat{a}\right), cn\right)$
	$\mathcal{S}\left(pred_m\left(a,m\right) \mid \widehat{pred}_m\left(m,\hat{a}\right), cn\right)$
	$\mathcal{S}\left(pred_m\left(m,a\right) \mid \widehat{pred}_m\left(\hat{a},m\right), cn\right)$
	$\mathcal{S}\left(pred_m\left(a,m\right) \mid \widehat{pred}_m\left(\hat{a},m\right), cn\right)$
	$\mathcal{S}\left(pred_m\left(m,*\right) \mid \widehat{pred}_m\left(m,*\right), cn\right)$
	$\vdots$

Predicate Schemas in Coreference

Predicate Schemas in Coreference

Pairwise Mention Scoring Function

$$f_{u,v} = \mathbf{w}^\top \phi(u, v)$$

Predicate Schemas in Coreference

Pairwise Mention Scoring Function

$$f_{u,v} = \mathbf{w}^\top \phi(u, v)$$

Scoring Function for Predicate Schemas

$$\mathbf{s}(u, v)$$

Predicate Schemas in Coreference

Pairwise Mention Scoring Function

$$f_{u,v} = \mathbf{w}^\top \phi(u, v)$$

Scoring Function for Predicate Schemas

$$\mathbf{s}(u, v)$$

We can add scores of Predicate Schemas as **Features**

Ways of Using Knowledge

- Major Disadvantages of Using Knowledge as **Features**
  - Noise in Knowledge
  - Inexplicit Textual Inference

- Alternative way
  - Using Knowledge as **Constraints**

Using Knowledge as Constraints

Using Knowledge as Constraints

Generating Constraints

$$\begin{cases} \text{if } s_i(u,v) \geq \alpha_i s_i(w,v) \Rightarrow y_{u,v} \geq y_{w,v}, \\ \text{if } s_i(u,v) \geq s_i(w,v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v} \end{cases}$$

Using Knowledge as Constraints

Generating Constraints

$$\begin{cases} \text{if } s_i(u,v) \geq \alpha_i s_i(w,v) \Rightarrow y_{u,v} \geq y_{w,v}, \\ \text{if } s_i(u,v) \geq s_i(w,v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v} \end{cases}$$

ILP inference (Best-Link)

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv}$$

$$\text{s.t. } \sum_{u<v} y_{uv} \leq 1, \forall v \qquad y_{uv} \in \{0,1\}$$

$$\text{if } s_{uv} \geq t + s_{um} \text{ then } y_{uv} \geq y_{um}$$

$$\text{if } s_{uv} \geq t' \cdot s_{um} \text{ then } y_{uv} \geq y_{um}$$

Scores for Predicate Schemas

- Multiple Sources
  - Gigaword
  - Wikipedia
  - Web Search
  - Polarity Information

Scores for Predicate Schemas

- Gigaword
  - Chunking + Dependency Parsing

    => predicate(subject, object)

    => Type 1 Predicate Schema
  - Heuristic Coreference

    => Type 2 Predicate Schema

Scores for Predicate Schemas

- Gigaword
  - Chunking + Dependency Parsing

    => predicate(subject, object)

    => Type 1 Predicate Schema
  - Heuristic Coreference

    => Type 2 Predicate Schema

Wikipedia

Entity Linking to ground on Wikipedia Entries (Disambiguation)

Gather Simple Statistics for 1) immediately after 2) immediately before 3) after 4) before

=> Type 1 Predicate Schema

) (approximation)

Scores for Predicate Schemas

- Web Search
  - Google Query with Quote (counts)
  - “m predicate”, “a m”, “a m”, “m predicate a”

    => Type 1 Predicate Schema

Scores for Predicate Schemas

- Web Search
  - ☐ Google Query with Quote (counts)
  - ☐ “m predicate”, “m a”, “a m”, “m predicate a”

    => Type 1 Predicate Schema

Polarity Information

Polarity on predicates => Polarity on mentions

Negate polarity if mention is object

Negate polarity for polarity-reversing connective

+1 if polarities for mentions are the same

-1 if polarities for mentions are different

=> Type 2 Predicate Schema

Recap

- Things to consider for using knowledge in NLP
  - Knowledge Representation
    - Predicate Schema
  - Knowledge Inference
    - Features VS. Inference
  - Knowledge Acquisition
    - Multiple Sources

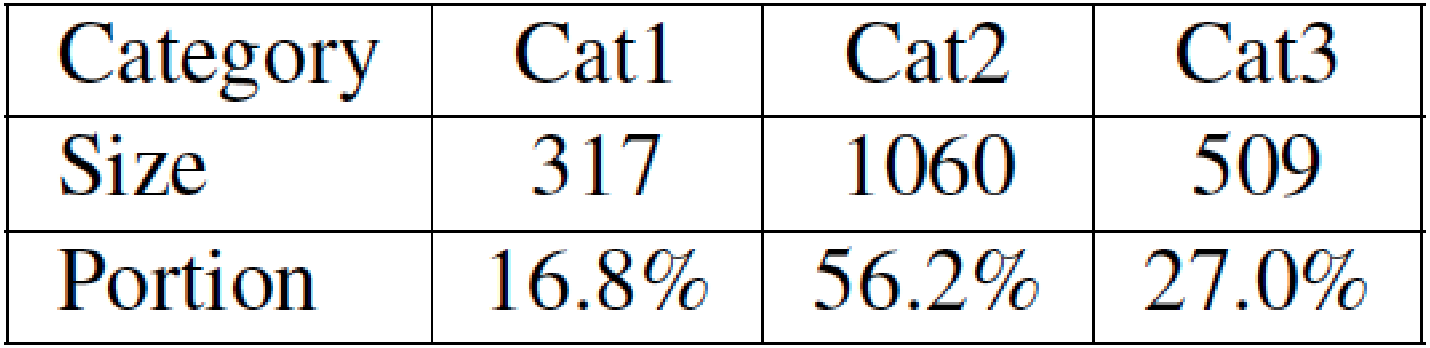
Datasets

- Winograd dataset[1]
  - [The bee] landed on [the flower] because [it] *had* pollen.
    [The bee] landed on [the flower] because [it] *wanted* pollen.

---

[1] http://www.hlt.utdallas.edu/~vince/data/emnlp12/

Datasets

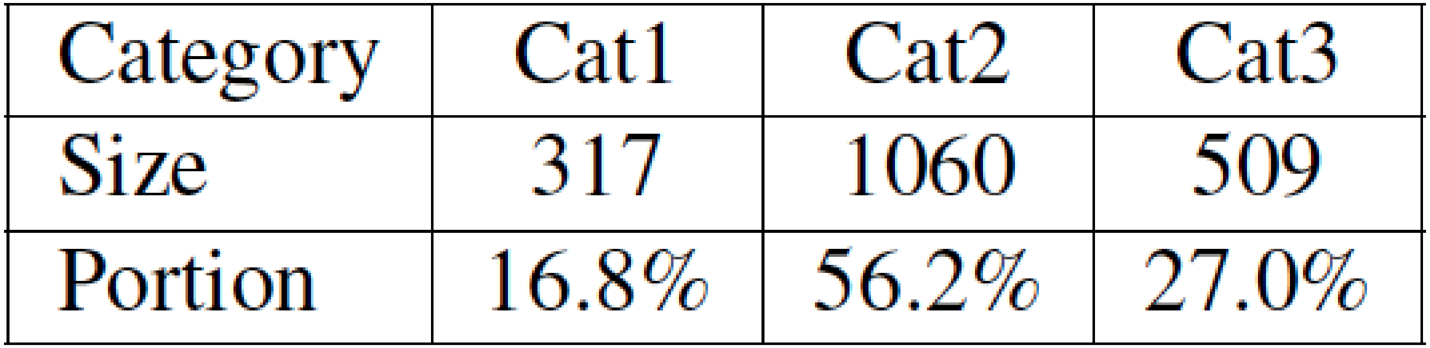
- Winograd dataset[1]
  - [The bee] landed on [the flower] because [it] *had* pollen.
    [The bee] landed on [the flower] because [it] *wanted* pollen.
  - 

Category	Cat1	Cat2	Cat3
Size	317	1060	509
Portion	16.8%	56.2%	27.0%

---

[1] http://www.hlt.utdallas.edu/~vince/data/emnlp12/

Datasets

- Winograd dataset[1]
  - [The bee] landed on [the flower] because [it] *had* pollen.
    [The bee] landed on [the flower] because [it] *wanted* pollen.
  - 

Category	Cat1	Cat2	Cat3
Size	317	1060	509
Portion	16.8%	56.2%	27.0%

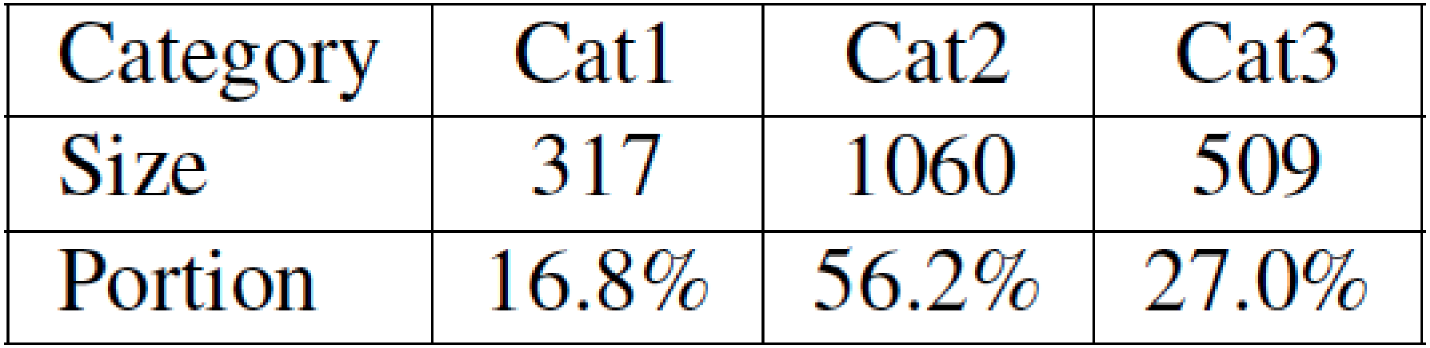
    [Jack] threw the bags of [John] into the water since [he] mistakenly asked **him** to carry **his** bags.

---

[1] http://www.hlt.utdallas.edu/~vince/data/emnlp12/

Datasets

- Winograd dataset[1]
  - ☐ [The bee] landed on [the flower] because [it] *had* pollen.
    [The bee] landed on [the flower] because [it] *wanted* pollen.
  - ☐



Category	Cat1	Cat2	Cat3
Size	317	1060	509
Portion	16.8%	56.2%	27.0%

[Jack] threw the bags of [John] into the water since [he] mistakenly asked **him** to carry **his** bags.

- WinoCoref dataset

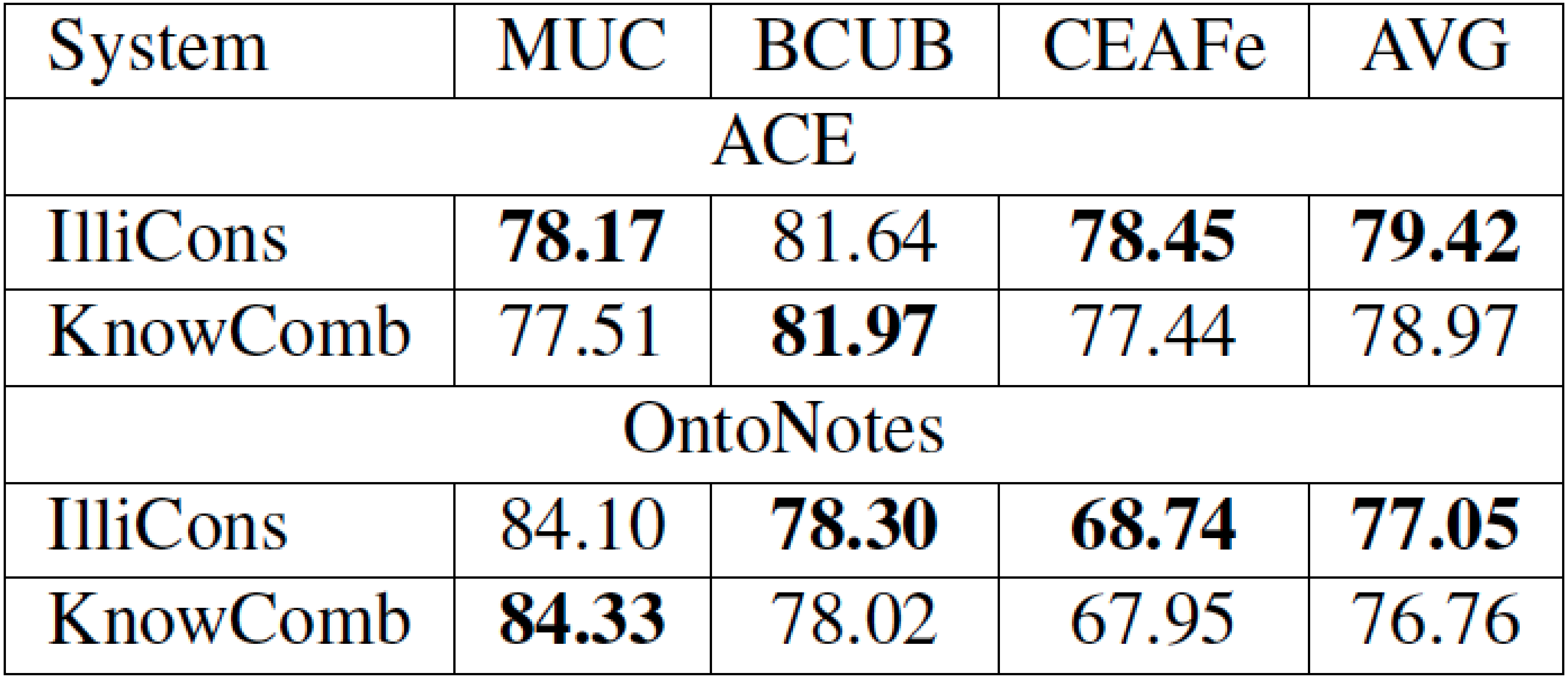
---

[1] http://www.hlt.utdallas.edu/~vince/data/emnlp12/

Evaluation on Hard Coreference

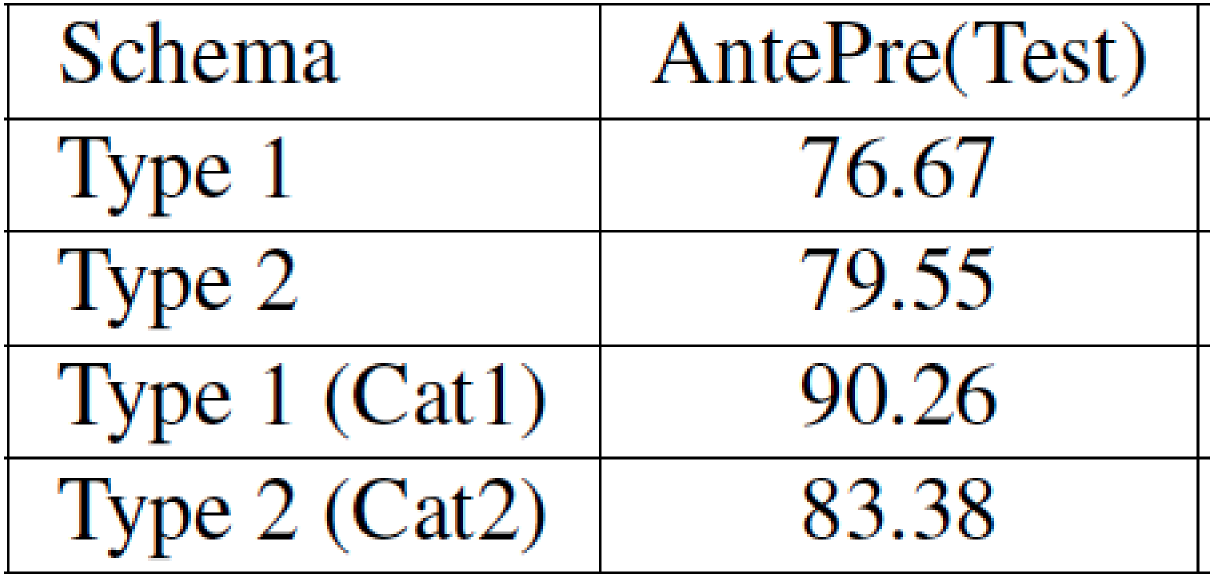
Dataset	Winograd	WinoCoref
Metric	Precision	AntePre
Illinois	51.48	68.37
IlliCons	53.26	74.32
Rahman and Ng (2012)	73.05	---
KnowFeat	71.81	88.48
KnowCons	74.93	88.95
KnowComb	**76.41**	**89.32**

Evaluation on Standard Coreference



System	MUC	BCUB	CEAFe	AVG
ACE				
IlliCons	**78.17**	81.64	**78.45**	**79.42**
KnowComb	77.51	**81.97**	77.44	78.97
OntoNotes				
IlliCons	84.10	**78.30**	**68.74**	**77.05**
KnowComb	**84.33**	78.02	67.95	76.76

Analysis on Effects of Schemas



Schema	AntePre(Test)
Type 1	76.67
Type 2	79.55
Type 1 (Cat1)	90.26
Type 2 (Cat2)	83.38

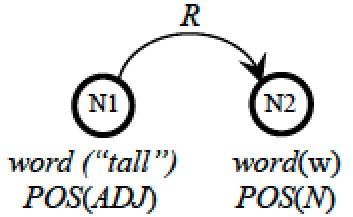
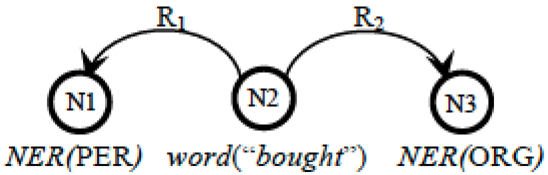
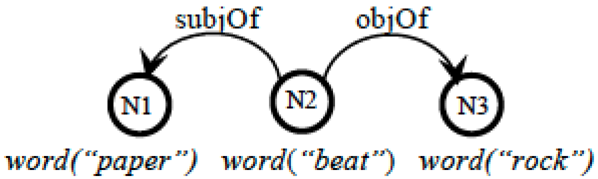
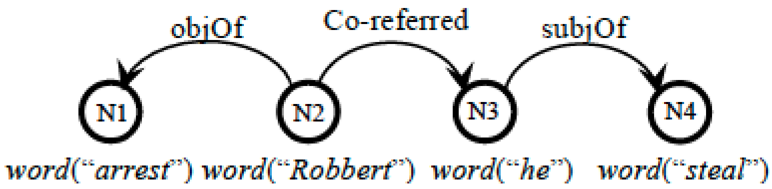
Part 2
Profiler: Knowledge Schemas at Scale

Goal

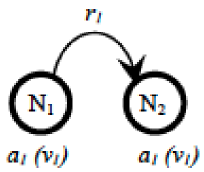
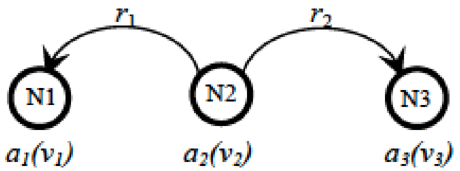
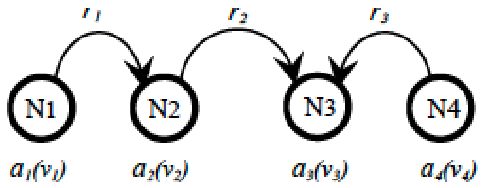
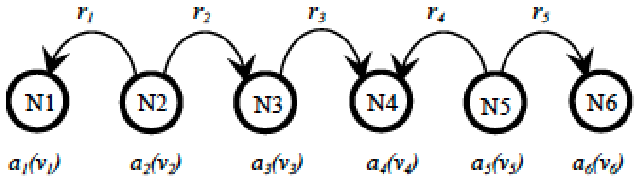
- How to enlarge the Knowledge acquired from text
  - Data Volume
  - Schema Richness

- Profiler
  - Demo: http://austen.cs.illinois.edu:60000/

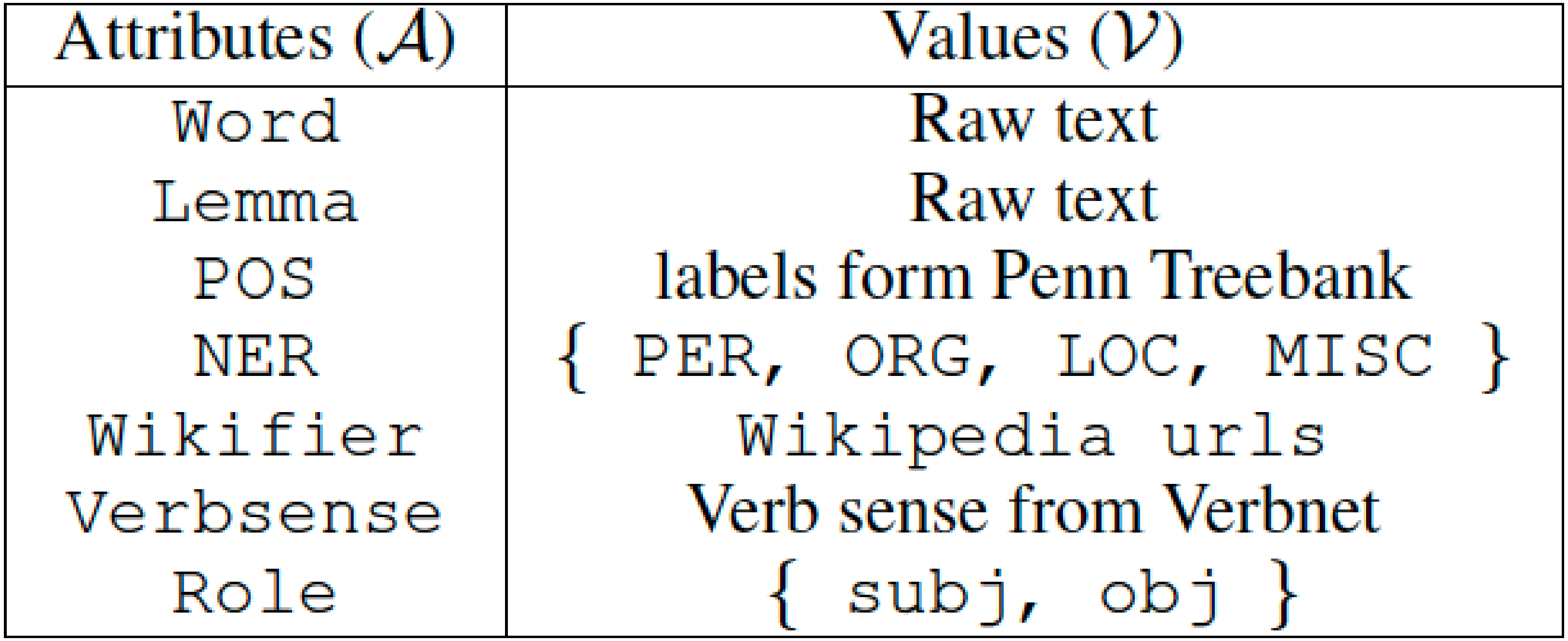
Motivation

#	Sentence	Schema Graph
1	"I chopped down [the tree] with my [axe] because [it] was tall."	
2	"[Larry Robbins], founder of Glenview Capital Management, bought shares of [Endo International Plc] ..."	
3	"Among [paper] and [rock] I chose rock, because [it] can beat scissors."	
4	"[Jimbo] attacked [Bobbert] because [he] stole an elephant from the zoo."	

Enriched Schemas

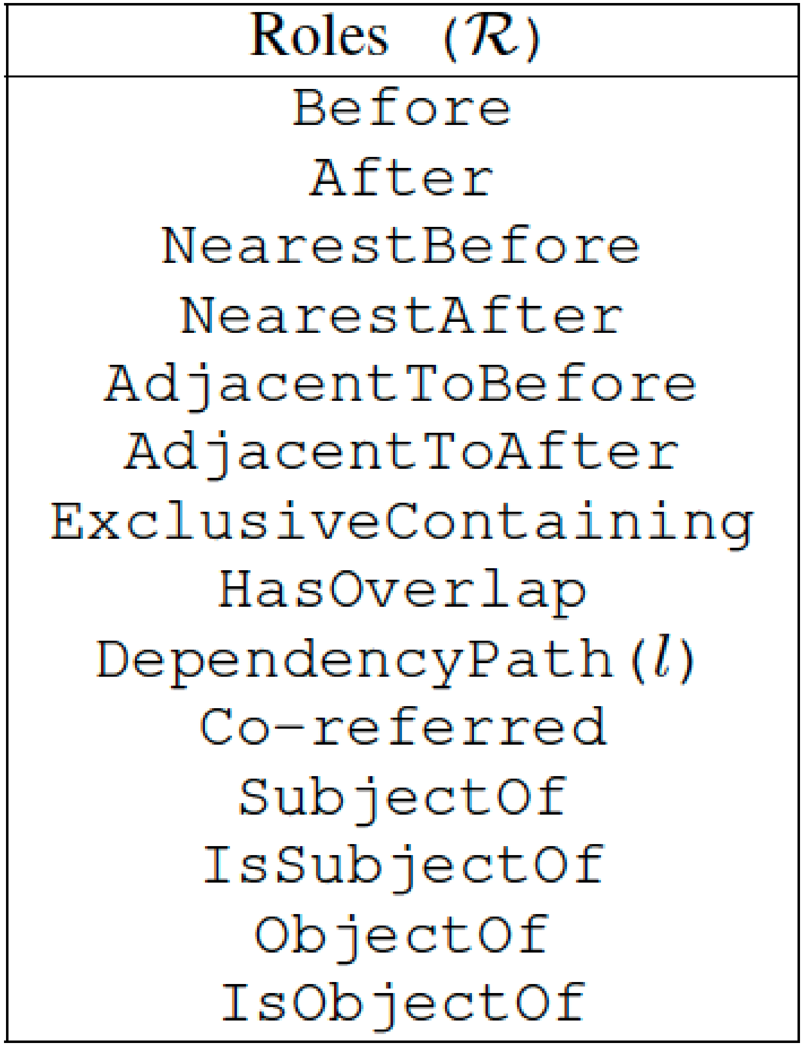
Concept Graph	Attributes = { Values }	Relations	Number of Schemas
	word = { set of words } POS = { Noun, Noun-Phrase, Verb, Verb-Phrase, Modifier } Wikifier = { URLS } Verbsense = { All verb senses }	Possible roles from Table 4 except Co-referred	24
	word = { set of words } POS = { Noun, Noun-Phrase, Verb, Verb-Phrase, Modifier } Verbsense ={ All verb senses }	Subj, ObjOf	2
	word = { All words } Verbsense = {All verb senses }	Subj, ObjOf, Co-referred	8
	word = { set of words } Verbsense = { All verb sense }	Subj, ObjOf, Co-referred	4

Enriched Schemas



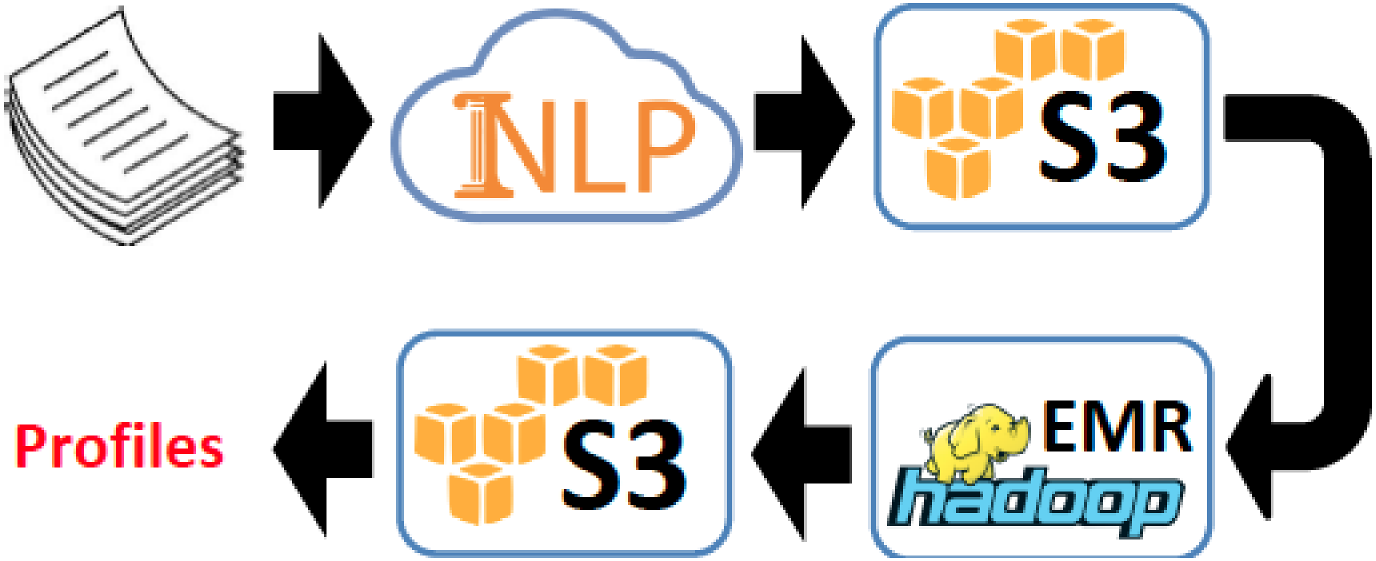
Attributes ($\mathcal{A}$)	Values ($\mathcal{V}$)
Word	Raw text
Lemma	Raw text
POS	labels form Penn Treebank
NER	{ PER, ORG, LOC, MISC }
Wikifier	Wikipedia urls
Verbsense	Verb sense from Verbnet
Role	{ subj, obj }

Enriched Schemas

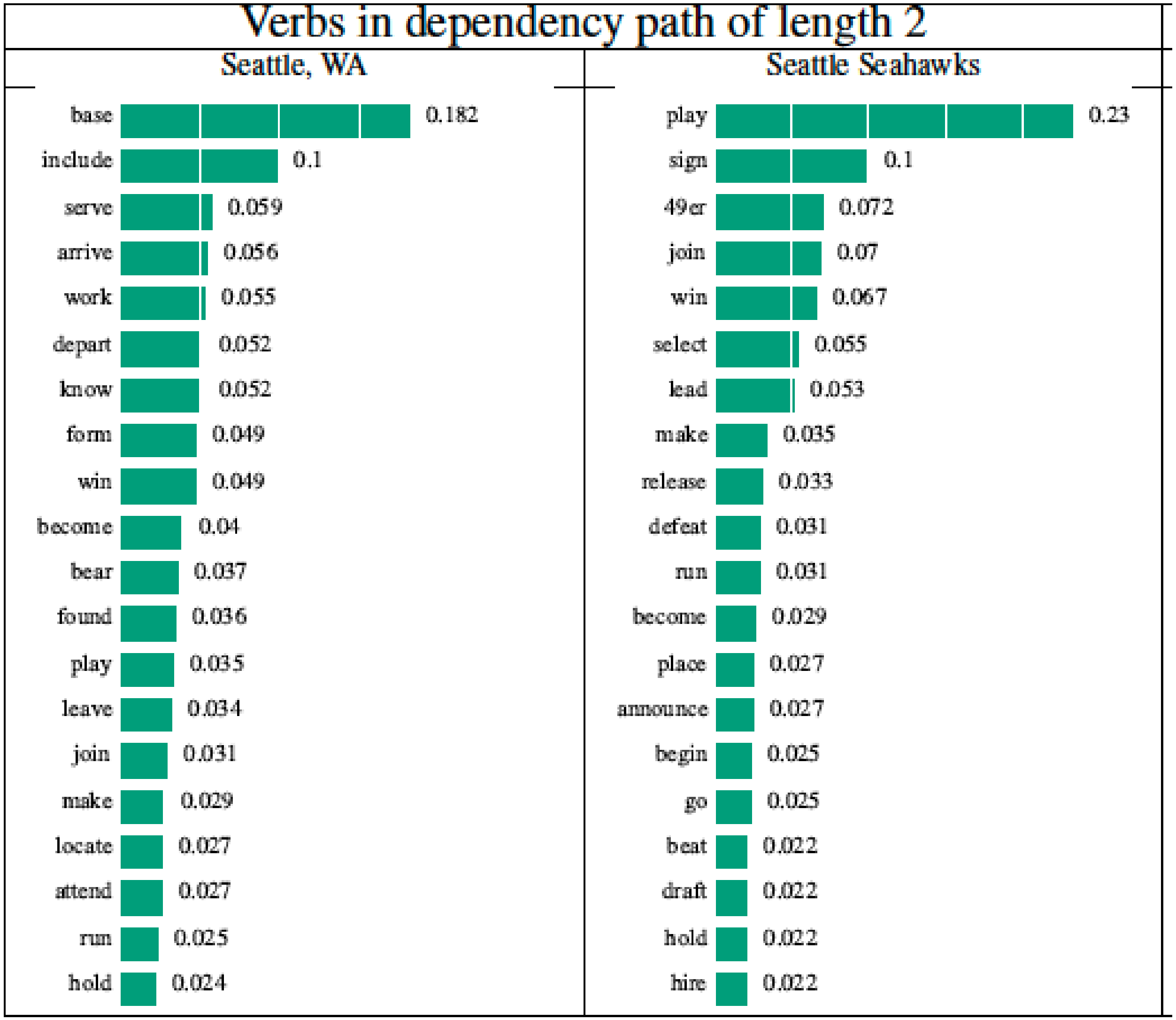


Roles ($\mathcal{R}$)
`Before`
`After`
`NearestBefore`
`NearestAfter`
`AdjacentToBefore`
`AdjacentToAfter`
`ExclusiveContaining`
`HasOverlap`
`DependencyPath`(l)
`Co-referred`
`SubjectOf`
`IsSubjectOf`
`ObjectOf`
`IsObjectOf`

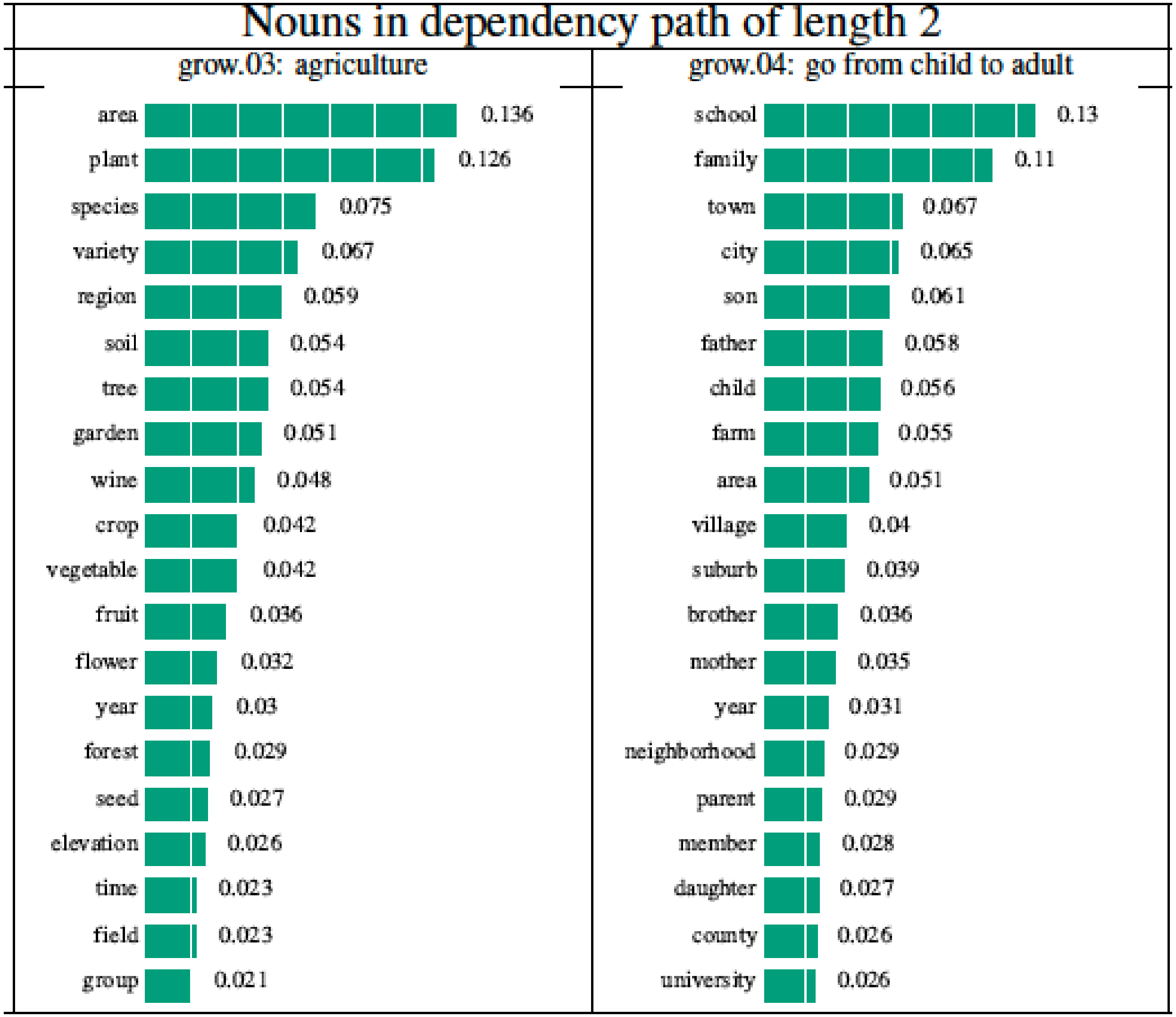
Implementation



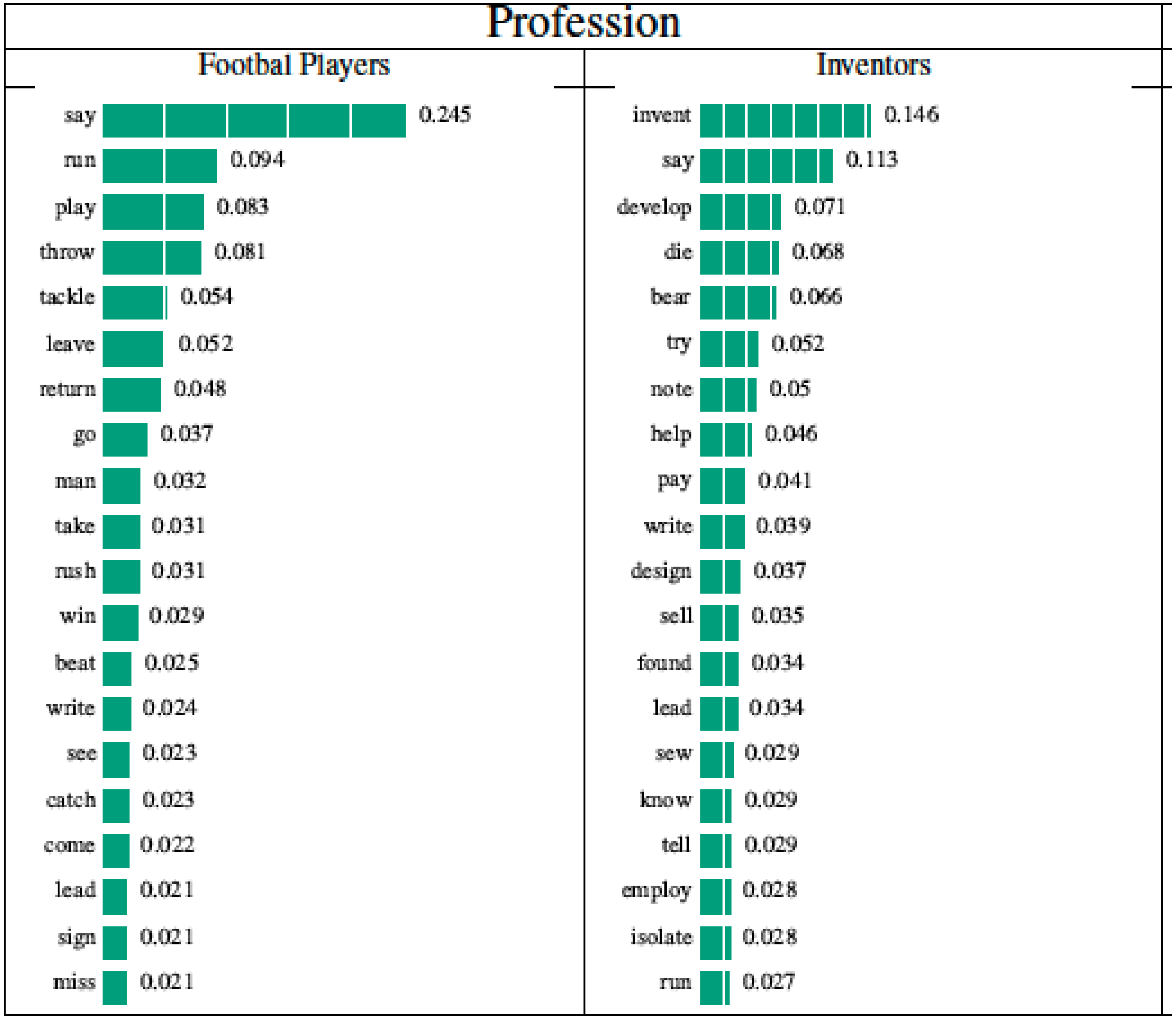
Effect of Wikification (Entity-Linking)



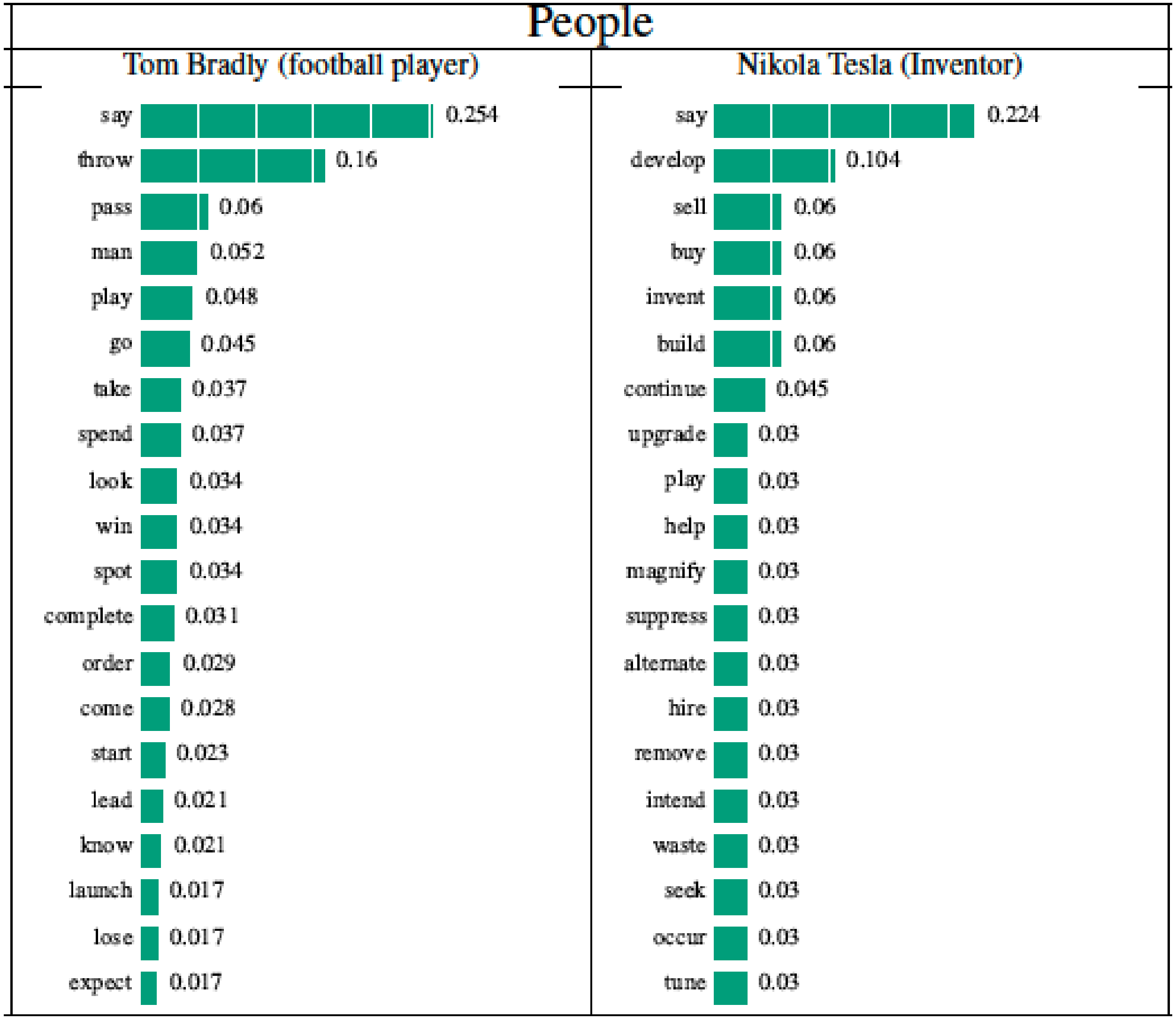
Effect of Wikification (Entity-Linking)



Knowledge Visualization



Knowledge Visualization



Publications

- [1] Solving Hard Coreference Problems. *Haoruo Peng****,** *Daniel Khashabi* and Dan Roth*. NAACL 2015.
- [2] A Joint Framework for Mention Head Detection and Coreference Resolution. Submitted to ACL 2015.
- [3] Profiler: Knowledge Schemas at Scale. Submitted to Transactions of ACL 2015 .

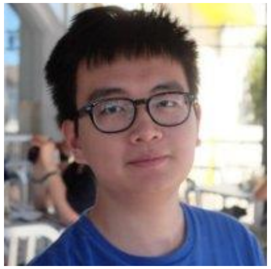
Future Directions

- The use of world knowledge in NLP tasks
  - Knowledge Representation (schemas)
    - Is co-occurrence information enough?
  - Knowledge Inference
    - Sparsity Issues
  - Knowledge Acquisition
    - Which sources to choose?
    - Interpolation /
  - Tasks beyond CR (CR can be seen as a subset of AI-complete problems)

- Outlier Detection for Singleton Mentions

Collaborators

Daniel Khashabi



Zhiye Fei

Kaiwei Chang

Prof. Dan Roth

Thank You !

Difficulties in CR

- Hard Coreference Problems
  - Ex.1 [A bird] perched on the [limb] and [it] bent.
  - Ex.2 [Robert] is robbed by [Kevin], and [he] is arrested by police.
  - Gender / Plurality information cannot help
  - -> Requires Knowledge

Difficulties in CR

- Hard Coreference Problems
  - Ex.1 [A bird] perched on the [limb] and [it] bent.
  - Ex.2 [Robert] is robbed by [Kevin], and [he] is arrested by police.
  - Gender / Plurality information cannot help
  - -> Requires Knowledge

Difficulties in CR

- Hard Coreference Problems
  - Ex.1 [A bird] perched on the [limb] and [it] bent.
  - Ex.2 [Robert] is robbed by [Kevin], and [he] is arrested by police.
  - Gender / Plurality information cannot help
  - -> Requires Knowledge

Difficulties in CR

- Hard Coreference Problems
  - ☐ Ex.1 [A bird] perched on the [limb] and [it] bent.
  - ☐ Ex.2 [Robert] is robbed by [Kevin], and [he] is arrested by police.
  - ☐ Gender / Plurality information cannot help
  - ☐ -> Requires Knowledge

Performance Gaps

System	Dataset	Gold	Predicted	Gap
Illinois	CoNLL-12	77.05	60.00	17.05
Berkeley	CoNLL-11	76.68	60.42	16.26
Stanford	ACE-04	81.05	70.33	10.72

-> Requires Better Mention Detection

A Joint Framework for Mention Head Detection and Coreference Resolution

- Goal: Improve CR on predicted mentions (End-to-End)
- Solution:

[Multinational companies investing in [China]] had become so angry that [they] recently set up an ***anti-piracy league*** to pressure [the [Chinese] government] to take action. [Domestic manufacturers, [who] are also suffering], launched a similar body this month.

  - ☐ Traditional: MD -> Coref
  - ☐ Our paper: Mention Head -> Joint Coref -> Head to Mention
  - ☐ Joint Learning / Inference Step
    - Add decision variables to decide whether to choose a head or not
    - Joint Coref is able to reject some mention head candidates
- Results

Dataset	Illinois	Baseline	Our Paper
ACE-04	68.27	68.27	71.20
CoNLL-12	60.00	61.71	63.01

ILP formulation of CR

- Best-Link with Knowledge Constraints

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv}$$

$$\text{s.t.} \quad \sum_{u<v} y_{uv} \leq 1, \; \forall v \qquad y_{uv} \in \{0,1\}$$

$$\text{if } s_{uv} \geq t + s_{um} \text{ then } y_{uv} \geq y_{um}$$

$$\text{if } s_{uv} \geq t' \cdot s_{um} \text{ then } y_{uv} \geq y_{um}$$

- Best-Link with Joint Mention Detection

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv} + \sum_m \lambda_m y_m$$

$$\text{s.t.} \quad \sum_{u<v} y_{uv} \leq 1, \; \forall v \qquad y_{uv} \in \{0,1\}$$

$$\sum_u y_{uv} \leq y_v, \; \forall v \qquad y_v \in \{0,1\}$$