

---

# **Overview of the TAC2010 Knowledge Base Population (KBP) Track**

---

Heng Ji, Ralph Grishman, Hoa Trang Dang and Kira Griffitt

November 15, 2010

---

# Outline

- Motivation of KBP Tasks
  - KBP2010 Task Overview
  - Participants Overview
  - Data Annotation and Analysis
  - Evaluation Metrics
  - Evaluation Results
  - Discussions and Lessons
-

---

# The Goal of KBP

- General Goal
    - Promote research in discovering facts about entities and expanding a knowledge source
  - What's New
    - Extraction at large scale (> 1 million documents) ;
    - Using a representative collection (not selected for relevance);
    - Cross-document entity resolution (extending the limited effort in ACE);
    - Linking the facts in text to a knowledge base;
    - Distant (and noisy) supervision through Infoboxes;
    - Rapid adaptation to new relations;
    - Support multi-lingual information fusion (KBP2011);
    - Capture temporal information (KBP2011)
-

---

# Knowledge Base Population (KBP2010) Task Overview

---

# KBP Setup

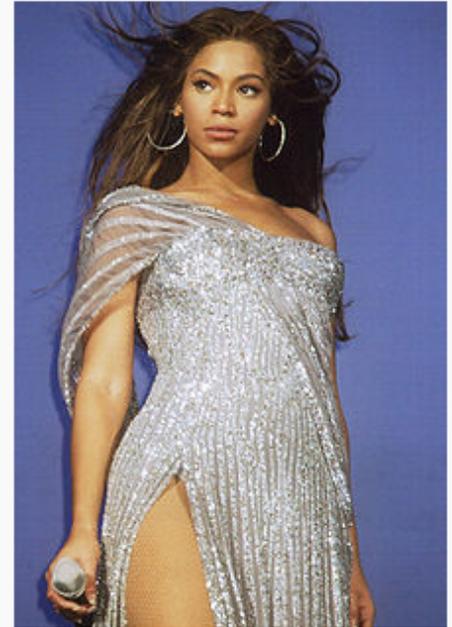
pop recording artist, actress and fashion model. Born and raised in Houston, Texas, she enrolled in various performing arts programs and was first exposed to singing and dancing competitions as a child. Knowles rose to fame in the late 1990s as the lead singer of the R&B girl group *Destiny's Child*, one of the world's best-selling girl groups of all time.

With *Destiny's Child*, Knowles released her debut solo album *Dangerously in Love* (2003), which spawned the number one single "Baby Boy" and became one of the most successful albums of that year, earning her a then record-tying five Grammy Awards. [3] Following the group's disbandment in 2005, Knowles released *B'Day* in 2006. It debuted at number one on the *Billboard* 200 and included the hits "Déjà Vu", "Irreplaceable" and "Beautiful Liar". Her third solo album *I Am... Sasha Fierce*, released in 2008, included the anthemic "Single Ladies (Put a Ring on It)". The album and its singles earned her six Grammy Awards, including the record for most Grammy Awards won by a female artist in one night. [4][5][6] Knowles is one of the most honored artists by the *Billboard* among female artists, with 16 awards—13 as a solo artist and three as a member of *Destiny's Child*. [7][8]

Knowles launched her acting career in 2001, appearing in the musical film *Carmen: A Hip Hopera*. In 2006, she starred in the lead role in the film *Dreamgirls*, for which she earned two Golden Globe nominations. Knowles launched her fashion line, *House of Deréon*, in 2004, and has endorsed such brands as *Pepsi*, *Tommy Hilfiger*, *Armani* and *L'Oréal*. In 2010, Knowles was ranked at number two on its list of the 100 Most Powerful and Influential Celebrities in the world, [9][10] she was also listed as the most powerful and influential musician in the world. [11] *Time* also included Knowles on its list of the "100 Most Influential People in the World". [12]

- Knowledge Base (KB)
  - Attributes (a.k.a., “slots”) derived from Wikipedia infoboxes are used to create the reference KB
- Source Collection
  - A large corpus of newswire and web documents (>1.3 million docs) is provided for systems to discover information to expand and populate KB

## Beyoncé Knowles



Knowles performing in 2007

### Background information

Birth name	Beyoncé Giselle Knowles
Born	September 4, 1981 (age 29) <sup>[1]</sup>
Origin	Houston, Texas, United States
Genres	R&B, pop, soul
Occupations	Singer-songwriter, record producer, actress, dancer, choreographer, model, video director
Instruments	Vocals
Years active	1997–present



# Entity Linking: Create Wiki Entry?

Shocking **Jim Parsons** truths revealed after Emmy win

August 29, 2010 | 7:22 pm



**Jim Parsons**

They reveal details after his Emmy win Sunday for lead actor in a comedy series. Interesting, or just odd? You decide.

For example: "I'm a big reader of almanacs, or I was, and I like lists -- oh, I sound a bit OCD, don't I?"

"The nerd will dance out," he says, when he mixes his character's "genius" thing with the writers' words.

And when he was very young, he played the kola kola bird in Rudyard Kipling's "The Elephant's Child," donning a pair of yellow tights



**NIL**

Query = "James Parsons"

---

# Entity Linking Task Definition

- Involve Three Entity Types
    - Person, Geo-political, Organization
  - Regular Entity Linking
    - Names must be aligned to entities in the KB; can use Wikipedia texts
  - Optional Entity linking
    - Without using Wikipedia texts, can use Infobox values
  - Query Example

```
<query id="EL000304">
  <name>Jim Parsons</name>
  <docid>eng-NG-31-100578-11879229</docid>
</query>
```
-

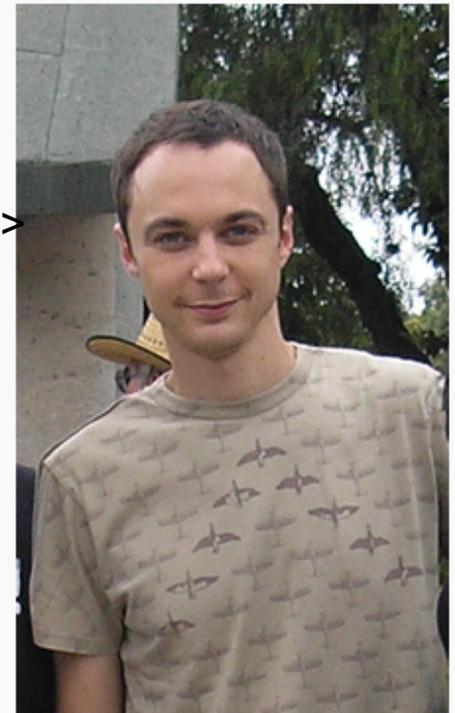
# Slot Filling: Create Wiki Infoboxes?



```
<query id="SF114">
  <name>Jim Parsons</name>
  <docid>eng-WL-11-174592-12943233</docid>
  <enttype>PER</enttype>
  <nodeid>E0300113</nodeid>
  <ignore>per:date_of_birth
    per:age per:country_of_birth
    per:city_of_birth</ignore>
</query>
```

Jim Parsons, a graduate of **School Attended: University of Houston** and Dance, won the Emmy on Sunday for Lead Actor in a Comedy Series for his work on The Big Bang Theory.

Jim Parsons



Parsons in 2008

<b>Born</b>	James Joseph Parsons March 24, 1973 (age 37) Houston, Texas, U.S.
<b>Occupation</b>	Actor
<b>Years active</b>	2000–present

# Regular Slot Filling

Person		Organization
per:alternate_names	per:title	org:alternate_names
per:date_of_birth	per:member_of	org:political/religious_affiliation
per:age	per:employee_of	org:top_members/employees
per:country_of_birth	per:religion	org:number_of_employees/members
per:stateorprovince_of_birth	per:spouse	org:members
per:city_of_birth	per:children	org:member_of
per:origin	per:parents	org:subsidiaries
per:date_of_death	per:siblings	org:parents
per:country_of_death	per:other_family	org:founded_by
per:stateorprovince_of_death	per:charges	org:founded
per:city_of_death		org:dissolved
per:cause_of_death		org:country_of_headquarters
per:countries_of_residence		org:stateorprovince_of_headquarters
per:stateorprovinces_of_residence		org:city_of_headquarters
per:cities_of_residence		org:shareholders
per:schools_attended		org:website

---

# Surprise Slot Filling

- Research on Portability of IE/QA techniques
    - how easily and rapidly an IE system can be adapted to new types of relations and events
  - Sites were given a maximum of 4 days to develop their systems
  - KBP2010 Surprise Slot Types
    - Per: diseases
    - Per: awards-won
    - Per: charity-Supported
    - Org: products
-

---

# Differences from KBP2009

- Entity Linking
    - Added Optional Entity Linking Task
  - Slot Filling
    - Location slots were each divided into three slots (city, state/province, country)
    - Origin was changed from a single to a list value
    - Dropped Geo-political Entities as a query type
    - Added Surprise Slot Filling Task
  - Query Selection
    - Independent selection processes for entity linking and slot filling
  - Submission Rules
    - Alternative runs were encouraged to evaluate the impact of web access and external knowledge base and offline Wikipedia text mining
    - At least one submission had to run without web access
-

---

# Participant Overview

---

# KBP2010 Participants

Team Name	Reg. EL	Opt. EL	Reg. SF	Surp. SF	Team Name	Reg. EL	Opt. EL	Reg. SF	Surp. SF
ARPANI	√				NUSchime	√			
BUDAPES T-ACAD	√	√	√		NYU			√	
BUPTPRIS	√		√		SIEL	√	√	√	
CMCRC	√				SMU	√			
CORTEX			√	√*	STANFORD			√	
CUNY	√		√	√	STANFORD _UBC	√			
HLTCOE	√	√	√		TCAR	√	√		
IBM			√		UBC			√	√
ICL	√	√	√		UC3M	√	√		
IIRG			√	√	USFD	√		√	
LCC	√		√	√	WebTLab	√			
LSV			√						

# KBP2010 Participants

- over 45 teams registered for KBP 2010 (not including the RTE-KBP Pilot task)
- Each team can submit up to 3 submissions

Tasks	2009		2010	
	#Teams	#Submissions	#Teams	#Submissions
Regular Entity Linking	13	35	16	46
Optional Entity Linking	-	-	7	20
Regular Slot Filling	8	16	15	31
Surprise Slot Filling	-	-	5	6

---

# Data Annotation and Analysis

---

# Data Annotation Overview

**Source collection: about 1.3 million newswire docs and 500K web docs, a few speech transcribed docs**

Entity Linking Corpus	Genre/Source	Size (entity mentions)		
		Person	Organization	GPE
Training	2009 Training	627	2710	567
	2010 Web data	500	500	500
Evaluation	Newswire	500	500	500
	Web data	250	250	250

Slot Filling Corpus	Task	Source	Size (entities)	
			Person	Organization
Training	Regular Task	2009 Evaluation	17	31
		<b>2010 Participants</b>	25	25
		2010 LDC	25	25
	Surprise Task	2010 LDC	16	16
Evaluation	Regular Task	LDC	50	50
	Surprise Task	LDC	20	20

---

# Data Selection

---

---

# Data Selection: Inputs

- KBP 2010 source data
    - KBP 2009 newswire data (~1M documents)
    - KBP 2010 weblog data (~500k documents)
  - Named Entity (NE) tagger output from NYU:
    - Namestrings selected from source data, labeled with type information (i.e. PER, ORG, or GPE)
    - Number of possible KB matches for a namestring
    - Number of KBP 2010 source data documents containing a namestring
-

---

# Data Selection: Entity Linking Stage 1

- Namestrings selected for the Entity-Linking task were confusable:
    - Quantitatively - No or many possible KB matches:
      - 0 or 7+ possible KB matches
    - Qualitatively – Entities difficult to determine:
      - Incorrect or alternative spelling (e.g. 'Jon' vs. 'John')
      - Abbreviated (e.g. 'CDC')
      - Ambiguous or more common name (e.g. 'John Smith')
-

---

## Data Selection: Entity Linking Stage 2

- 20 documents were selected from the source data for each of the Entity Linking namestrings
    - If more than 20 documents were found, the selected set equally represented the unique entities possibly referred to by the namestring
  - Example: 'John W' namestring
    - Source data search produced 50 documents containing 'John W.'
    - In these documents, three unique entities were referred to by the namestring 'John W.' (e.g. John Williams, John Wagner, John Wilson)
    - The 20 documents selected would have equally represented these three unique entities
-

---

# Data Selection: Slot-Filling

## Stage 1

Namestrings selected for the Slot-Filling task:

- Were not confusable:
    - Quantitatively - Few possible KB matches
      - 1-6 possible KB matches
    - Qualitatively - Easy to determine entity namestrings:
      - Correct spelling (e.g. 'Bill Clinton')
      - Not abbreviated (e.g. 'American Medical Association')
      - Unambiguous or less common name (e.g. 'Tony Blair')
  - Had slot-filling answers in source data
  - Had an incomplete or non-existent KB entry
-

---

# Data Selection: Slot-Filling

## Stage 2

- Searched source data to ensure there were slot-filling answers for each selected namestring
    - 2-3 slot-filling answers were needed for a SF entity to remain in pipeline
  - Also searched source data for a reference document for each SF entity
    - Selection of the reference document created namestring-document pair, which was used to determine if SF entity had a KB entry
    - If no reference document was available, the namestring was excluded from the pipeline
-

---

## Data Selection:

### Excluded Namestrings

- Excluded Namestrings did not meet:
    - The confusability requirements for the Entity Linking or Slot Filling tasks
      - Namestrings had to meet both the quantitative AND qualitative criteria
    - General data-selection standards
      - incorrectly typed by NE tagger
      - Nonsensical
      - objectionable content
-

---

# KB-Linking:

Stage 3 for both EL and SF entities

- Searched to determine whether entities in namestring-document pairs had KB entries
    - If it was unclear whether an entity had a KB entry, it was removed from the pipeline
    - Entity Linking entities marked “KB entry” or “No KB entry” were released as Entity Linking queries
    - Slot-Filling entities marked “KB entry” or “No KB entry” were reviewed and used in slot-filling annotation
-

---

# Slot-Filling Annotation

---

---

# Slot-Filling Annotation Preparations

- KB entries for all SF entities were reviewed to prevent redundant annotations.
    - Already-filled slots were made visible to annotators but could not be altered.
  - Slot-Filling Guidelines Developed
    - Used 2009 Assessment guidelines as a basis
    - Added descriptions of slot-filling task and annotation categories
  - In-House Annotator Training
    - Done in-house so that potential annotators could complete a practice kit with supervision.
-

---

## Slot-Filling: Annotation

- Annotators were presented with an entity and all available slots for that entity
    - If an entity was NOT linked to the KB, annotators were presented with
      - All single-value slots
      - All list-value slots
    - If an entity was linked to the KB, annotators were presented with
      - Empty single-value slots
      - All list-value slots
  - Slot-filling annotation performed within a two-hour time limit for each SF entity
-

---

# Slot-Filling Assessment

---

## Updates to Assessment Guidelines

- In the interim between Slot-Filling annotation and Assessment, continued discussions between LDC staff and TAC coordinators on questions raised by annotators produced clarifications on slot descriptions
    - Classification of specific GPEs (Emirates, Capital Districts)
    - Employers of actors
    - Residence of Universities
    - Headquarters of sports teams
    - Non-recognized states and areas of contention
    - Alternate names
-

---

# Slot-Filling: Assessment

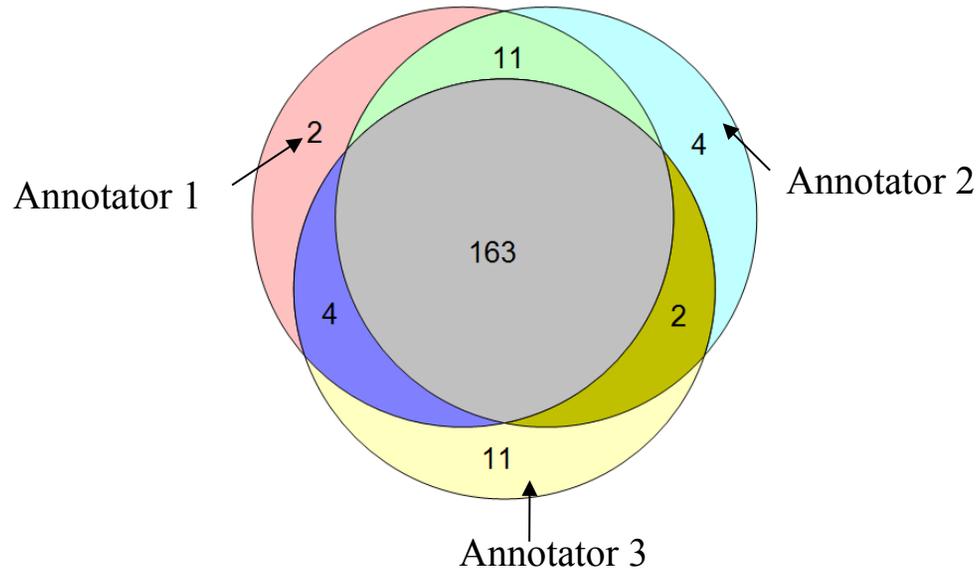
- Potential assessors were rigorously tested
    - Required to complete a full-size assessment kit
      - 12 slots filled with slot-filling answers for an actual SF entity
    - 90% accuracy or greater required to pass
      - A number of candidates did not make the cut
    - Questions raised during the test resulted in further clarifications to the guidelines:
      - Acceptability of past and future relationships
      - PER:Age, PER:Alternate Names, PER:Country of Birth, PER:Employee Of, PER:Origin
      - ORG:Alternate Names and ORG:Number of Employees/Members
  - Those who passed went on to assess the validity of slot-filling answers from both humans and machines
-

---

# Slot-Filling: Assessment

- **Quality Control**
    - After assessment was completed, systematic and spot-checking QC was performed on 50% of the assessment data.
    - These QC passes revealed only incidental errors, which were corrected prior to release of assessment results
  - **Errors identified in assessment results will be incorporated into the 2011 slot-filling annotation and assessment guidelines**
-

# Entity Linking Inter-Annotator Agreement



Entity Type	#Total Queries	Agreement Rate	Genre	#Disagreed Queries
Person	59	91.53%	Newswire	4
			Web Text	1
Geo-political	64	87.5%	Newswire	3
			Web Text	5
Organization	57	92.98%	Newswire	3
			Web Text	1

# Slot Filling Human Annotation Performance

## ■ Evaluation assessment of LDC Hand Annotation

Performance	P(%)	R(%)	F(%)
All Slots	70.14	54.06	61.06
All except per:top-employee, per:member_of, per:title	71.63	57.6	63.86

## ■ Why is the precision only 70%?

- 32 responses were judged as inexact and 200 as wrong answers
- A third annotator's assessment on 20 answers marked as wrong:  
65% incorrect; 15% correct; 20% uncertain
- Some annotated answers are not explicitly stated in the document
  - ... some require a little world knowledge and reasoning
- Ambiguities and underspecification in the annotation guideline
- Confusion about acceptable answers
- Updates to KBP2010 annotation guideline for assessment

---

# Evaluation Metrics

---

## Entity Linking Scoring Metric

- Micro-averaged Accuracy (official metric)
  - Mean accuracy across all queries
- Macro-averaged Accuracy
  - Mean accuracy across all KB entries

---

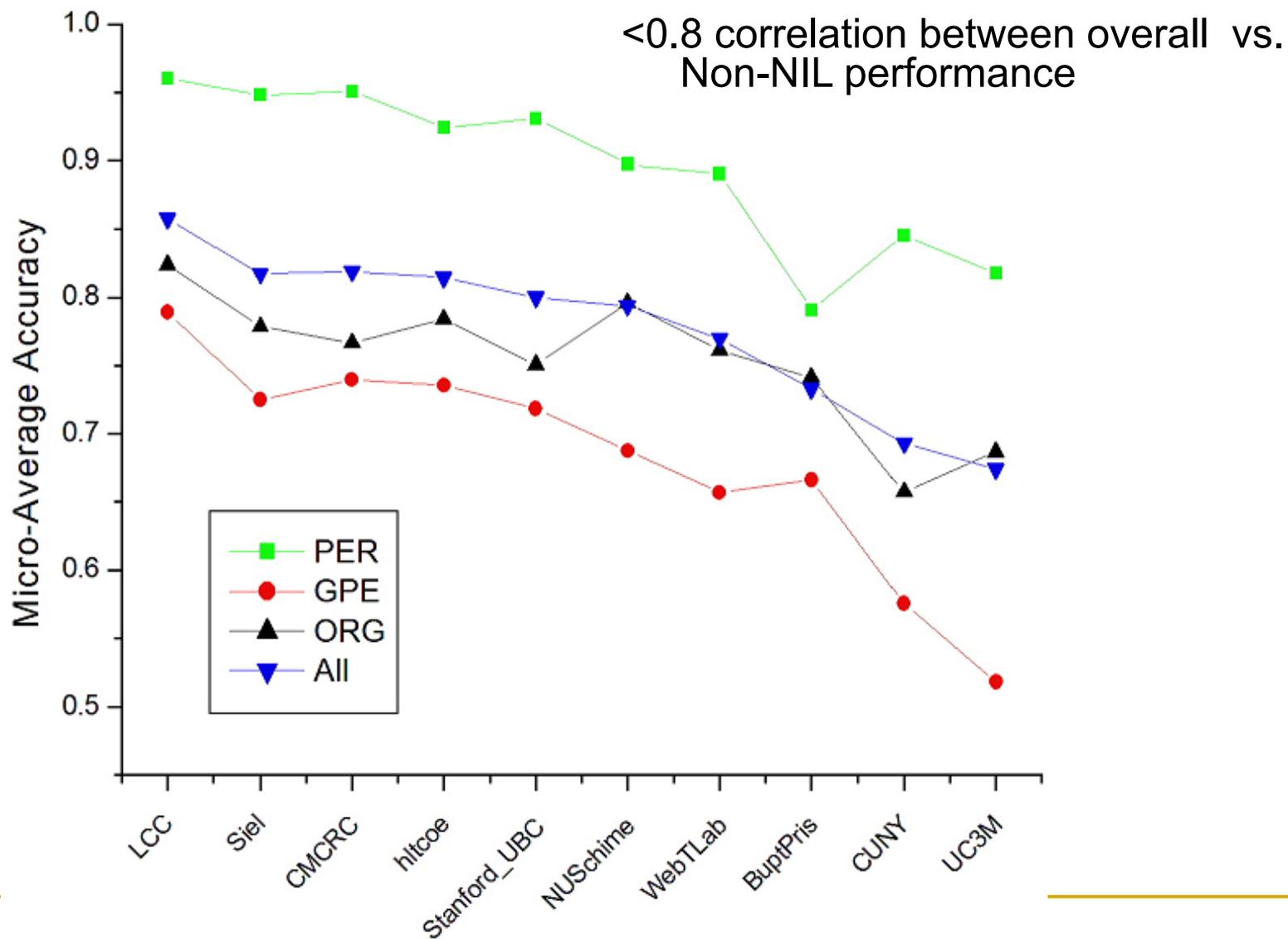
## Slot Filling Scoring Metric

- Each response is rated as correct, inexact, redundant, or wrong (credit only given for correct responses)
  - Redundancy: (1) response vs. KB; (2) among responses: build *equivalence class*, credit only for one member of each class
- Correct = # (non-NIL system output slots judged correct)
- System = # (non-NIL system output slots)
- Reference =
  - # (single-valued slots with a correct non-NIL response) +
  - # (equivalence classes for all list-valued slots)
- Standard Precision, Recall, F-measure

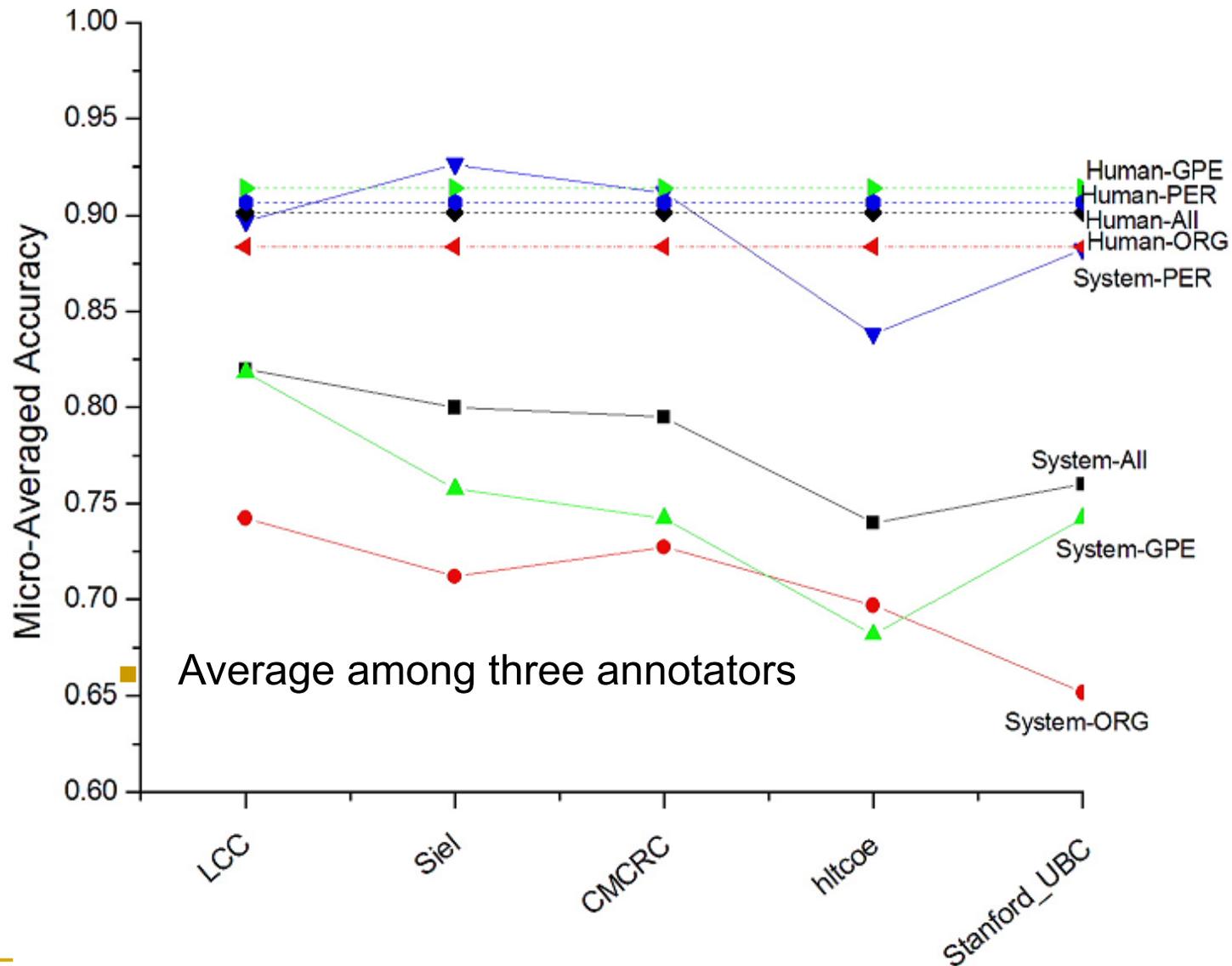
---

# Evaluation Results

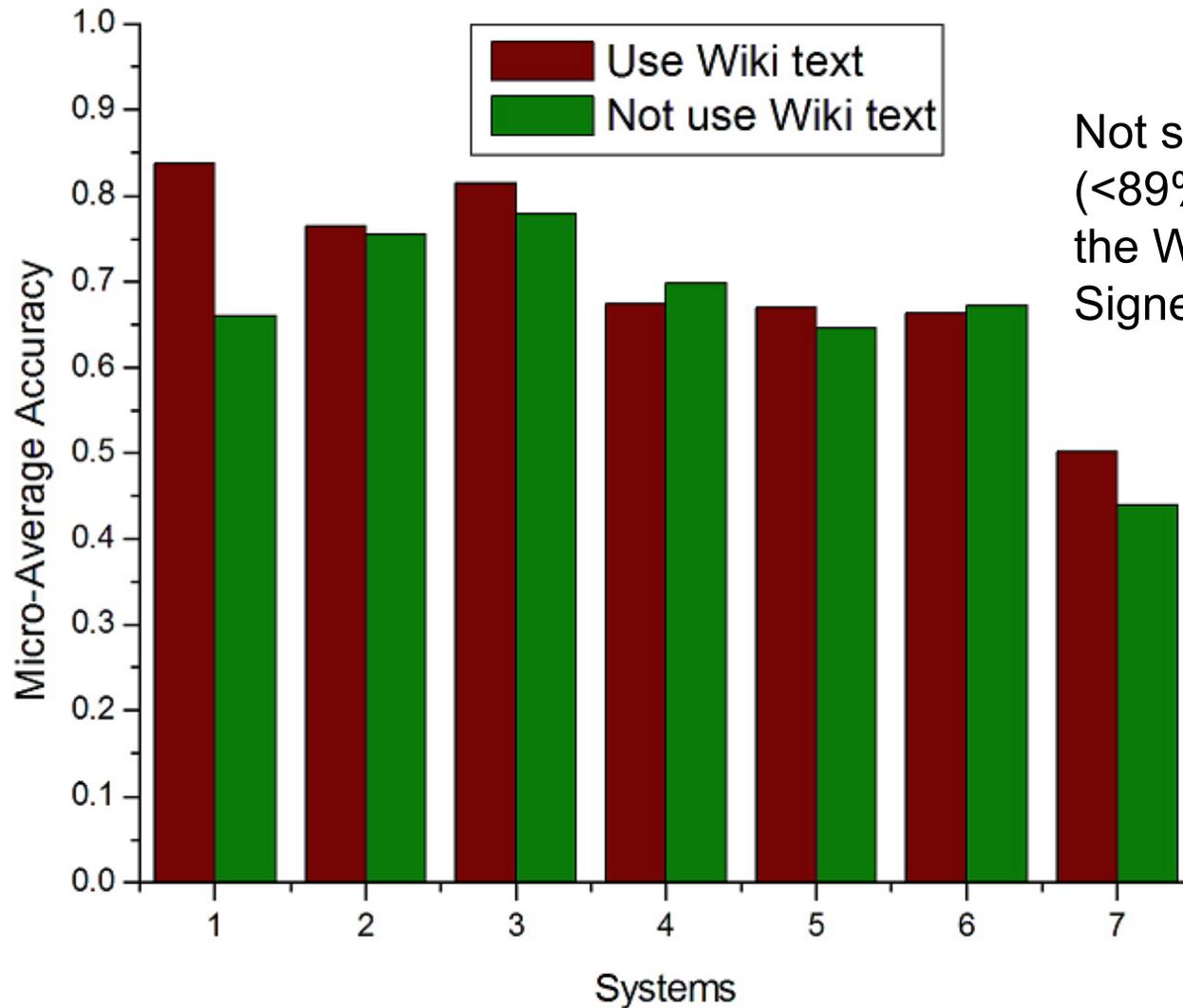
# Top-10 Regular Entity Linking Systems



## Human/System Entity Linking Comparison (subset of 200 queries)

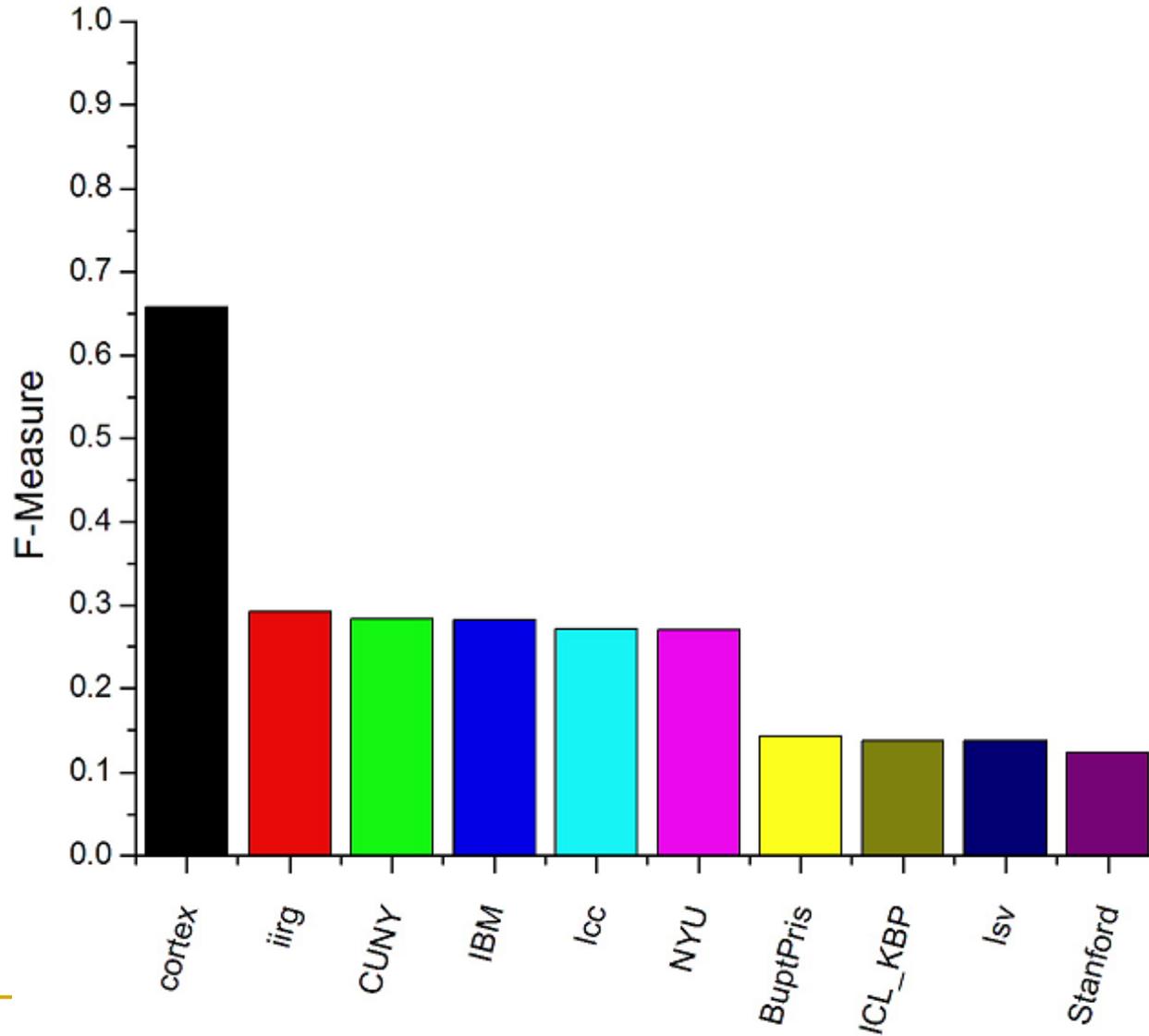


# Impact of Wikipedia Texts on Entity Linking



Not statistically significant  
( $<89\%$  confidence level using  
the Wilcoxon Matched-Pairs  
Signed-Ranks Test)

# Top-10 Regular Slot Filling Systems



## Surprise Slot Filling Evaluation Results

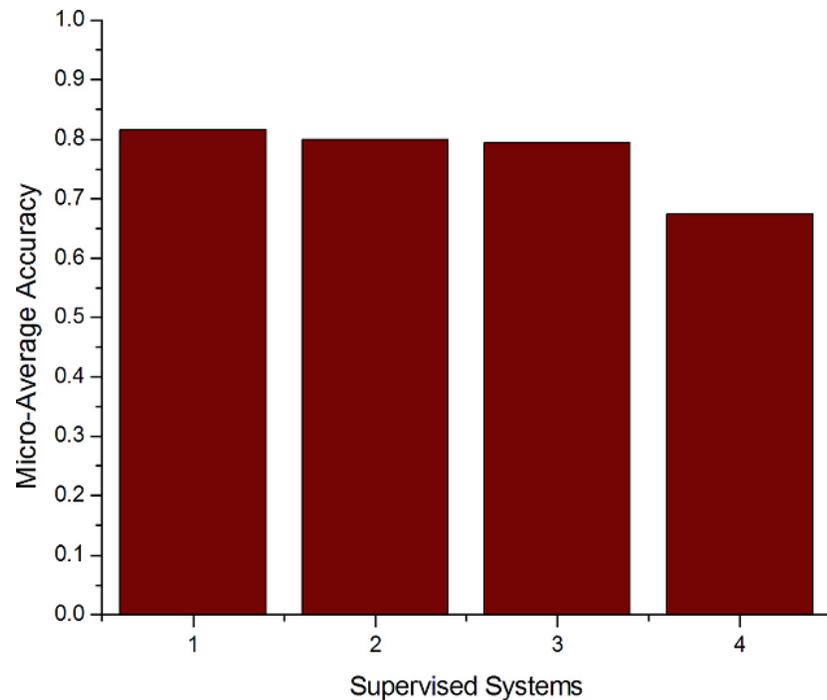
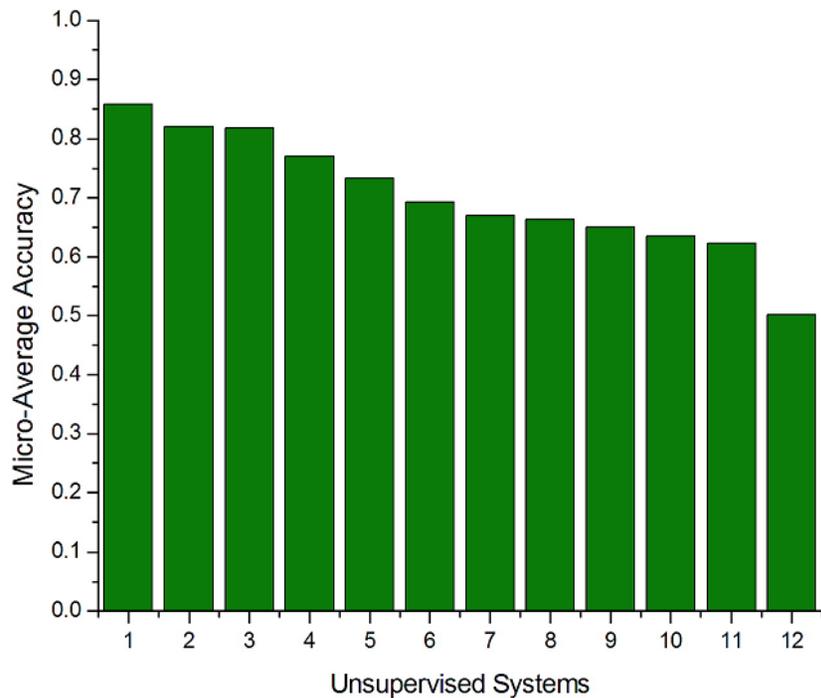
- Only 5 sites participated in the task, possibly because it was scheduled immediately after the regular task
- 4 sites fielded automatic systems
- Only one site, LCC, exceeded 10% F score

Time (#hours)	P(%)	R(%)	F(%)
11	50.33	15.45	23.64
34	52.36	24.16	33.06

---

# Discussion of Entity Linking

# Unsupervised/Minimally-Supervised vs. Supervised Learning for Entity Linking



# Impact of Semantic Features on Entity Linking

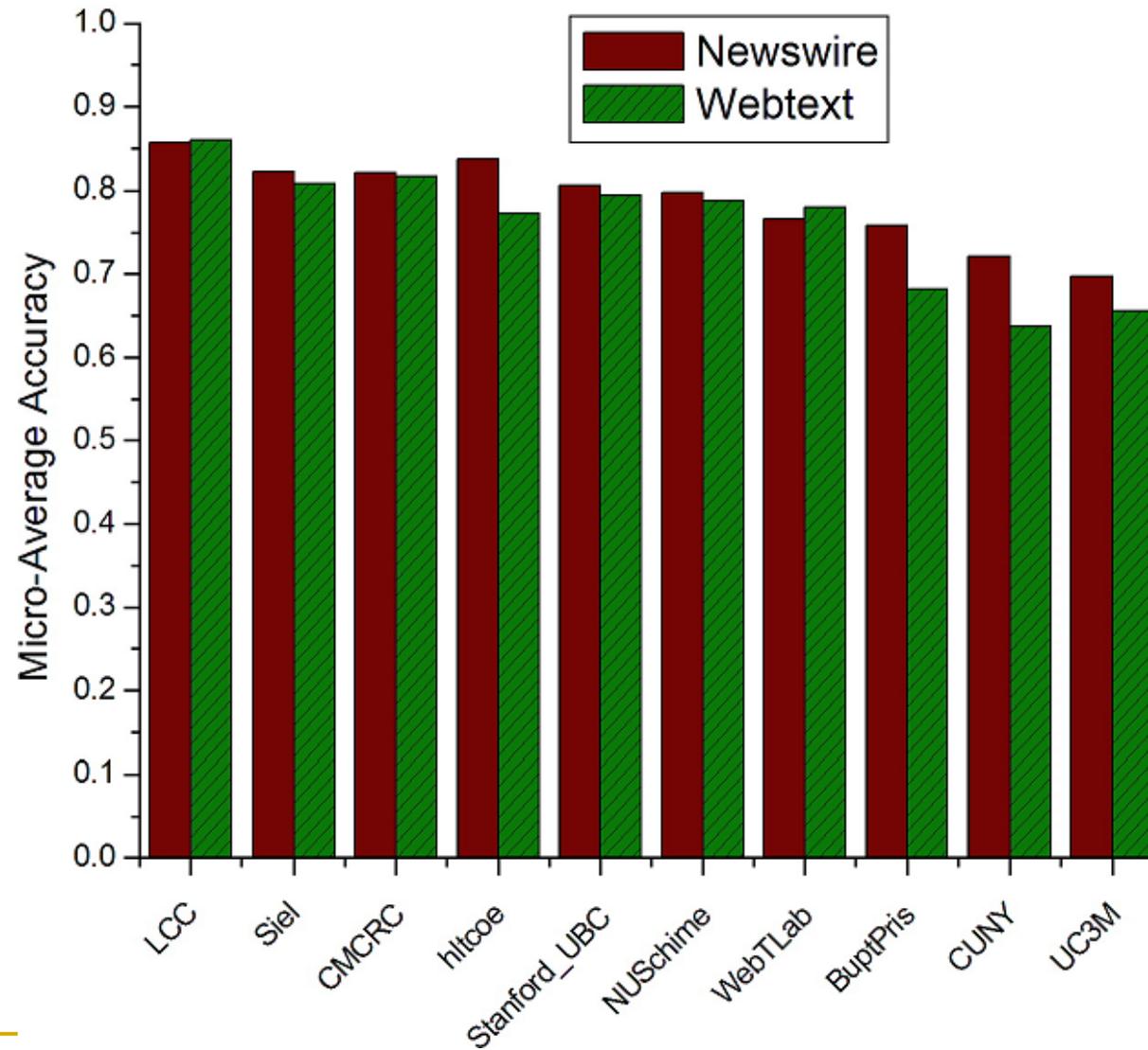
- CUNY-BLENDER: Use Slot Filling results as features (entity profile)

System	Person	Organization	Geo-Political	Overall
Without SF feedback	84.6%	63.1%	57.5%	59.9%
With SF feedback	92.8%	65.7%	84.1%	69.3%

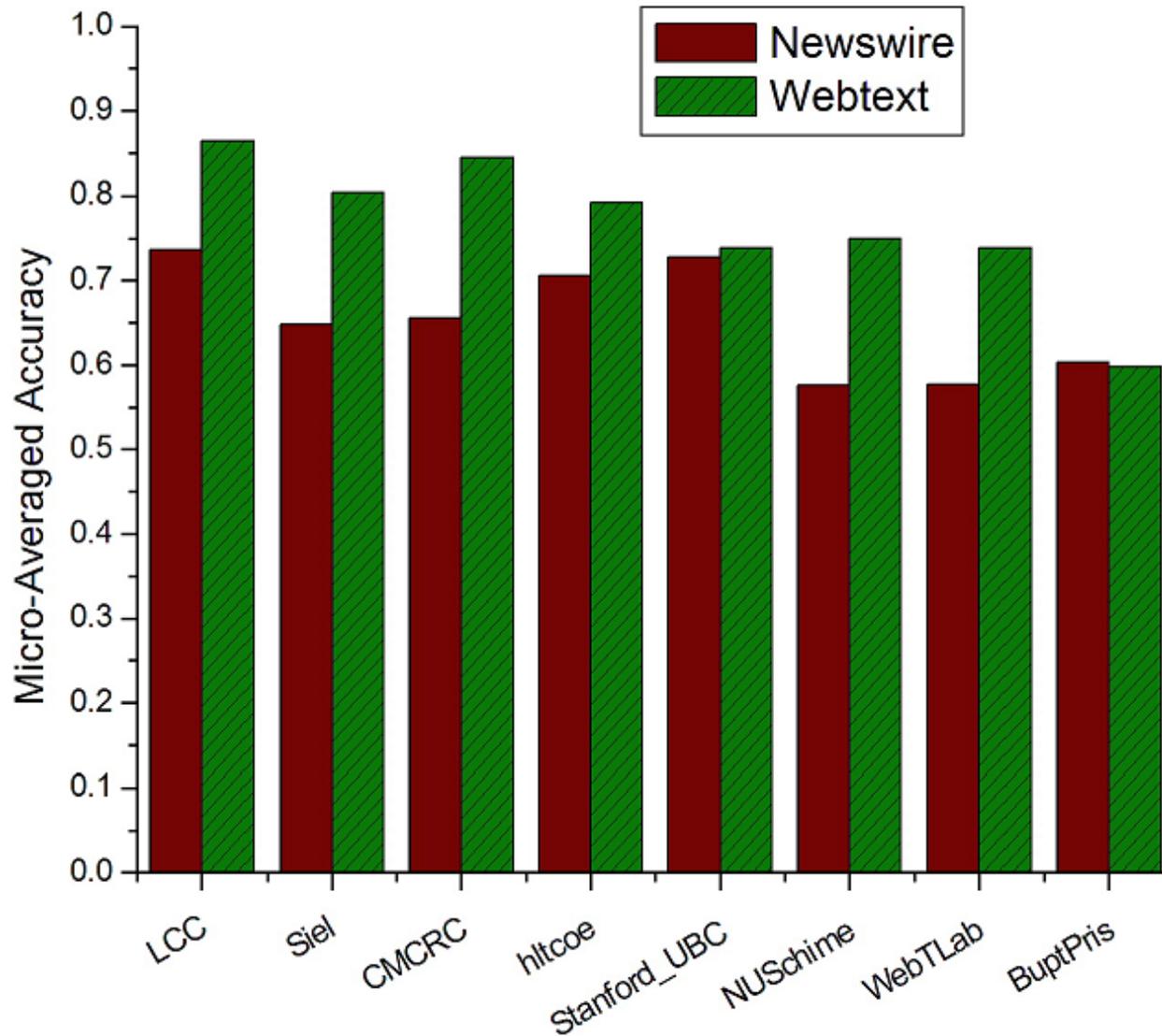
- BuptPris: Use name tagging, infoboxes etc. as features

System	Person	Organization	Geo-Political	Overall
Without semantic Features	83.9%	59.5%	33.4%	58.9%
With semantic features	79.1%	74.1%	66.6%	73.3%

## Impact of Data Genre (All queries)



## Impact of Data Genre (Non-NIL queries)



---

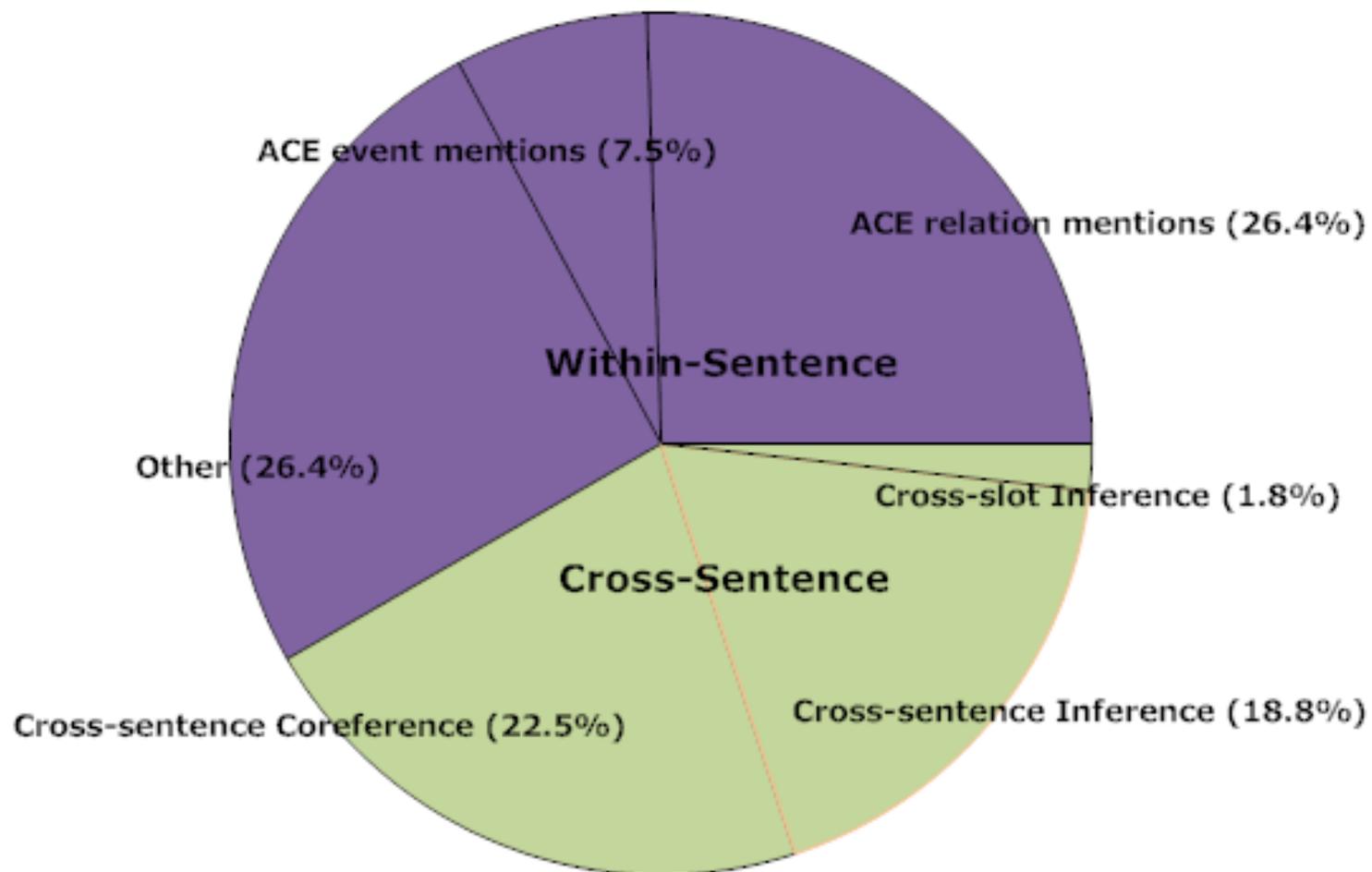
# Discussion of Slot Filling

---

# Slot-Specific Analysis

- A few slots account for a large fraction of the answers:
  - per:title, per:employee\_of, per:member\_of, and org:top\_members/employees account for 37% of correct responses
- For a few slots, delimiting exact answer is difficult ... result is 'inexact' slot fills
  - per:charges, per:title (“rookie driver”; “record producer”)
- For a few slots, equivalent-answer detection is important to avoid redundant answers
  - per:title again accounts for the largest number of cases. e.g., “defense minister” and “defense chief” are equivalent.

# How much Inference is Needed?



---

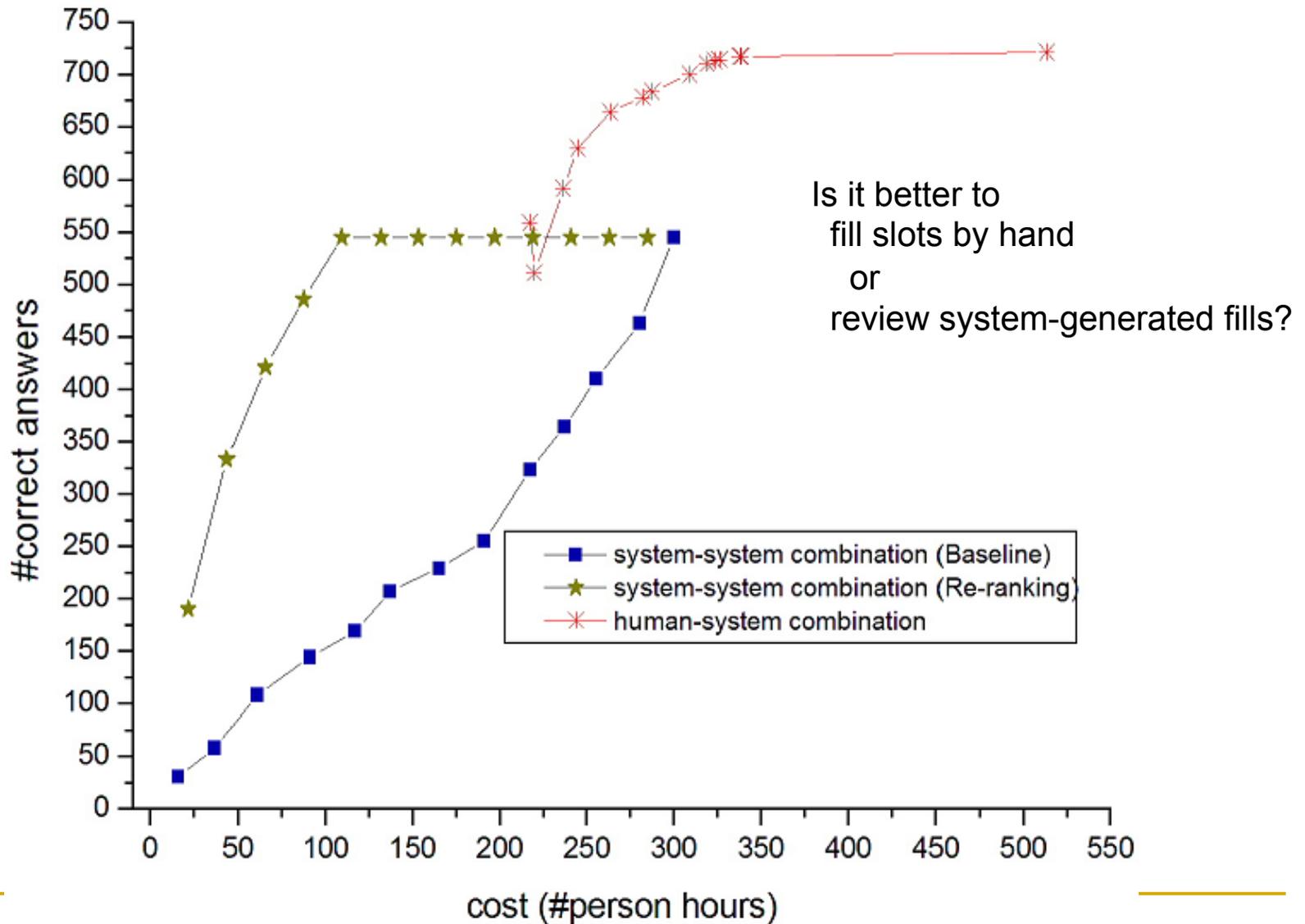
# Why KBP is more difficult than ACE

- Cross-sentence Inference – non-identity coreference(per:children)
  - *Lahoud* is married to an Armenian and the couple have three children. Eldest son *Emile Emile Lahoud* was a member of parliament between 2000 and 2005.
- Cross-slot Inference (per:children)
  - *People Magazine* has confirmed that actress *Julia Roberts* has given birth to her third child a boy named *Henry Daniel Moder*. Henry was born Monday in Los Angeles and weighed 8? lbs. Roberts, 39, and husband *Danny Moder*, 38, are already parents to twins *Hazel* and *Phinnaeus* who were born in November 2006.

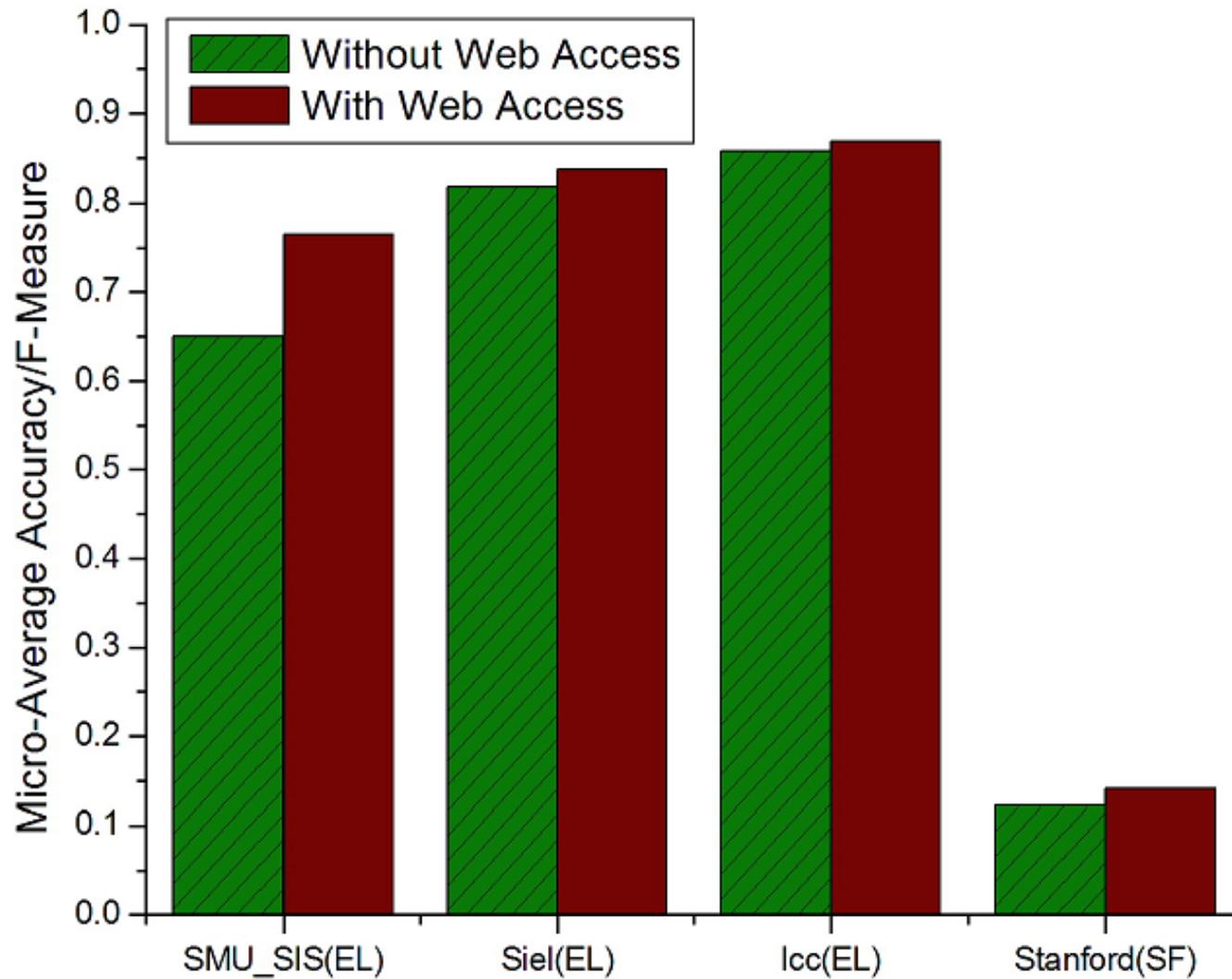
## Impact of Using External Knowledge/Wiki Text Mining for Answer Validation

System	Use Answer Validation?	Precision	Recall	F-measure
LCC	No	45.33	18.76	26.54
	Yes	44.87	19.44	27.13
CUNY	No	27.99	26.02	26.97
	Yes	28.74	27.85	28.29

# Answer-Key Preparation



## Impact of Web Access



---

# Conclusions

- KBP2010 was marked by a growing variety of tasks and a growing pool of participants
- Entity Linking
  - General improvement in performance over last year's results
  - system performance approached and in some cases (and for some entity types) exceeded the human benchmark
  - Performance on new web data was close to newswire
  - Optional task was slightly but not significantly worse
- Slot Filling
  - A wide variety of approaches were represented
  - Substantial training corpora produced
  - A better understanding of the shortcomings of these approaches will lead to continued progress