# Overview of the TAC 2009 Knowledge Base Population Track

16 November 2009

Paul McNamee (JHU/HLTCOE)
Heather Simpson (LDC)
Hoa Trang Dang (NIST)

# Talk Outline

- Background
- Task Description
- Data
- Target Selection
- Assessment
- Results
- Conclusion

# Motivation

- IE & QA technologies have been studied in isolation
  - Not focused on discovery of information <u>for inclusion in an existing knowledge base</u>
  - No consideration of novelty, contradiction

- Issues when filling in a KB
  - Accurate extraction of facts
  - Global resolution of entities
  - Maintaining provenance of asserted facts
  - Avoiding contradiction / detection of novel information
  - Temporal qualification of assertions
  - Leveraging existing KB to assist with extraction
  - Scalability

# Comparison to ACE & TREC-QA

- Corpus vs. document focus
  - ACE: component tasks (NER, relation extraction) for a set of isolated documents
  - KBP: learn facts from a corpus. Repetition not very important. Asserting wrong information is bad.
- Context
  - In KBP, there is a reference knowledge base, so avoiding redundancy and detecting contradiction are important
  - In KBP slots are fixed and targets change. In TREC QA, the targets dictated which questions were asked.
- Knowing when you don't know
  - TREC QA had a small percentage of NIL questions (4-10%)

# Participating Teams

| | |
|---|---|
| BUAP_1 | B. Autonomous University of Puebla |
| CSLU.OHSU | Oregon Health and Science University |
| DAMSEL | Macquarie University |
| HLTCOE | JHU Human Language Technology Center of Excellence |
| IBM | TJ Watson IBM Research |
| Janya | Janya Inc. |
| NLPR_KBP | National Laboratory of Pattern Recognition, China |
| PRIS | Beijing University of Posts and Telecommunications |
| QUANTA | Tsinghua University |
| Siel_09 | International Institute of Information Technology |
| Stanford_UBC | Stanford University |
| TCAR_r6a | National Security Agency |
| UC3M | Universidad Carlos III de Madrid |

# KBP Snapshot

- Track structure
  - NIST – overall organization, infrastructure, evaluation
  - LDC – develop and distribute data resources, target selection, human assessments
- Datasets
  - LDC produced 1.3M English newswire collection
  - Reference KB populated with semi-structured facts obtained from English Wikipedia (Oct '08 dump)
    - 200k people, 200k GPEs, 60k orgs, 300+k misc/non-entities
- Two tasks
  - Entity Linking - Grounding entity mentions in documents to KB entries
  - Slot Filling - Learning attributes about target entities

# Sample KB Entry

```
<entity wiki_title="Michael_Phelps”
        type="PER”
        id="E0318992”
        name="Michael Phelps">
<facts class="Infobox Swimmer">
<fact name="swimmername">Michael Phelps</fact>
<fact name="fullname">Michael Fred Phelps</fact>
<fact name="nicknames">The Baltimore Bullet</fact>
<fact name="nationality”>United States</fact>
<fact name="strokes”>Butterfly, Individual Medley, Freestyle, Backst
<fact name="club">Club Wolverine, University of Michigan</fact>
<fact name="birthdate">June 30, 1985 (1985-06-30) (age 23)</fact>
<fact name="birthplace”>Baltimore, Maryland, United States</fact>
<fact name="height">6 ft 4 in (1.93 m)</fact>
<fact name="weight">200 pounds (91 kg)</fact>
</facts>
<wiki_text><![CDATA[Michael Phelps
Michael Fred Phelps (born June 30, 1985) is an American swimmer. H
Olympic gold medals, the most by any Olympian. As of August 2008,
world records in swimming. Phelps holds the record for the most gol
single Olympics with the eight golds he won at the 2008 Olympic Gan
```


**Michael Phelps**

Michael Phelps at the 2008 Beijing Olympics

| Personal information | |
|---|---|
| Full name: | Michael Fred Phelps |
| Nickname(s): | The Baltimore Bullet[1] |
| Nationality: | United States |
| Stroke(s): | Butterfly, Individual Medley, Freestyle, Backstroke |
| Club: | Club Wolverine, University of Michigan |
| Date of birth: | June 30, 1985 (age 23) |
| Place of birth: | Baltimore, Maryland, United States |
| Height: | 6 ft 4 in (1.93 m) |
| Weight: | 200 pounds (91 kg) |
| Medal record [show] | |

# Most Frequent KB Classes

95142 settlement
72992 album
34659 film
32464 musical artist
23138 actor
21195 single
16765 company
15644 book
14567 football biography
14121 person
12646 radio station
12514 nrhp
12324 vg
11813 planet
10818 uk place
10113 television

8353 ort in deutschland
8061 university
7675 airport
7492 military person
7270 road
7185 indian jurisdiction
7123 cityit
6143 australian place
6131 mountain
5957 military conflict
5952 military unit
5937 city
5630 software
5501 mlb retired
5397 writer
5349 scientist

5222 lake
4913 television episode
4636 school
4426 commune de france
4265 aircraft
4229 ice hockey player
3918 german location
3234 nflactive
3168 disease
3070 politician
3036 u.s. county
2956 station
2950 automobile
2933 officeholder
2833 broadcast
2728 swiss town

PER ORG GPE OTHER

# Entity Linking Task

### John Williams

Richard Kaufman goes a long way back with **John Williams**. Trained as a classical violinist, Californian Kaufman started doing session work in the Hollywood studios in the 1970s. One of his movies was Jaws, with **Williams** conducting his score in recording sessions in 1975...

### Michael Phelps

Debbie Phelps, the mother of swimming star **Michael Phelps**, who won a record eight gold medals in Beijing, is the author of a new memoir, ...

**Michael Phelps** is the scientist most often identified as the inventor of PET, a technique that permits the imaging of biological processes in the organ systems of living individuals. **Phelps** has ...

| | | |
|---|---|---|
| John Williams | author | 1922-1994 |
| J. Lloyd Williams | botanist | 1854-1945 |
| John Williams | politician | 1955- |
| John J. Williams | US Senator | 1904-1988 |
| John Williams | Archbishop | 1582-1650 |
| John Williams | composer | 1932- |
| Jonathan Williams | poet | 1929- |

| | | |
|---|---|---|
| Michael Phelps | swimmer | 1985- |
| Michael Phelps | biophysicist | 1939- |

**Identify matching entry, or determine that entity is missing from KB**

# Related Work (1)

- **Cluster Documents Mentioning Entities**
  - **Mann & Yarowsky (CoNLL 2003)**
    - **Clustering with TFIDF/BoW+NNPs (F=77%) with the additional use of relation features (F=86%)**
  - **Gooi & Allan (HLT 2004)**
    - **Agglomerative Clustering (F=80%)**
  - **Studied at Web People Search workshops (WePS-1,2)**

- **Cross-Document Entity Coreference**
  - **Group together mentions of the same named entity across documents in a large corpus**
  - **Studied at ACE 2008 (English and Arabic)**

# Related Work (2)

- Add missing links between Wikipedia pages
  - Adafre and de Rijke (2005), Milne & Witten (2008), Fader et al. (2009)
  - Differences with KBP 2009
    - Include non-entities
    - Ignore NIL entities (those not in WP)
    - Cast problem as WSD

- Link entities to matching Wikipedia article
  - Bunescu & Pasca (2006) - Personal names (WP text)
  - Cucerzan (2007) - All entities (news articles, WP text)
  - Differences with KBP 2009
    - Ignore NIL entities
    - KBP worked with PER/ORG/GPEs; did not focus on popular entities

# Slot Filling Task

**Target: EPA**
**(plus 1 document)**

**Environmental Protection Agency**

| Agency overview | |
|---|---|
| Employees | 17,964 (2005) |
| Annual budget | $7.3 billion (2007) |
| Agency executive | Lisa P. Jackson, Administrator |

**Generic Entity Classes**

**Person, Organization, GPE**

**Missing information to mine from text:**

- **Date formed: 12/2/1970**
- **Website: http://www.epa.gov/**
- **Headquarters: Washington, DC**
- **Nicknames: EPA, USEPA**
- **Type: federal agency**
- **Address: 1200 Pennsylvania Avenue NW**

**Optional: Also want to link some learned values within the KB:**

- **Headquarters: Washington, DC (kbid: 735)**

# Entity Attributes

| Person | Organization | Geo-Political Entity |
| --- | --- | --- |
| alternate names | alternate names | alternate names |
| age | political/religious affiliation | capital |
| birth: date, place | top members/employees | subsidiary orgs |
| death: date, place, cause | number of employees | top employees |
| national origin | members | political parties |
| residences | member of | established |
| spouse | subsidiaries | population |
| children | parents | currency |
| parents | founded by | |
| siblings | founded | |
| other family | dissolved | |
| schools attended | headquarters | |
| job title | shareholders | |
| employee-of | website | |
| member-of | | |
| religion | | |
| criminal charges | | |

# Introduction

- Planned resources
  - Source data
  - Knowledge Base
  - Entity Linking and Slot Filling lists
  - System assessment
- Data Distribution
  - 291 copies of 9 unique corpora, to 31 individual organizations
    - Distributed under evaluation license which gives no cost access for purposes of TAC
  - Corpora will be published in LDC catalog

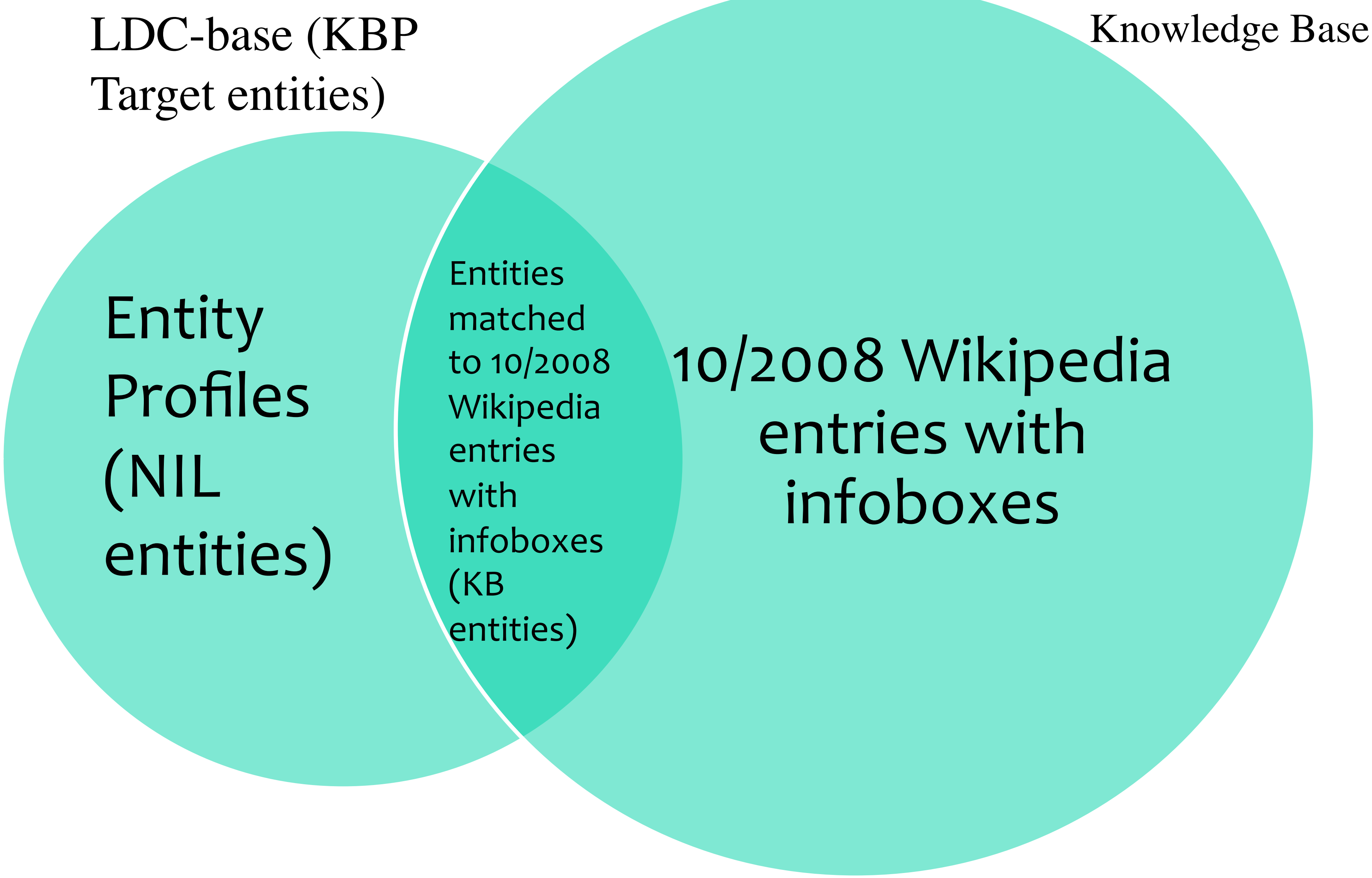
# Source Data Profile

- Volume
  - 1289649 documents, 6.5 GB
- Epoch
  - >99% from 2007 and 2008, to approximate epoch of the KB (10/2008)
  - 1994-07 through 2008-12 (ACE08 Evaluation docs)
- Genre
  - newswire
  - broadcast news and conversation
  - weblogs and newsgroups
- Selection
  - 10,000 previously unreleased documents selected for ACE08 Evaluation
  - added NW (2007-2008) from English Gigaword 4 (LDC2009T13)
- Processing
  - source files processed to ACE source document format (SGML)
  - parseable as XML

# Knowledge Base Description

- Based on October 2008 snapshot of Wikipedia
- Parsed into XML format from raw wiki markup
  - Only includes pages with (parseable) Infoboxes
    - Infobox fields parsed into <fact> elements
- Infoboxes standardized
  - NIST , LDC, JHU collaborated on Generic Infobox slots for Person, Organization, and GPE entity types
  - LDC created partial mapping from existing infobox types in KB to generic set
- LDC-Base vs. Knowledge Base
  - Knowledge Base - the XML data extracted from Wikipedia, distributed to the KBP teams
  - LDC-Base - LDC's internal database of entity information developed during annotation; used to produce materials for use in the project (e.g., the entity list, etc.)

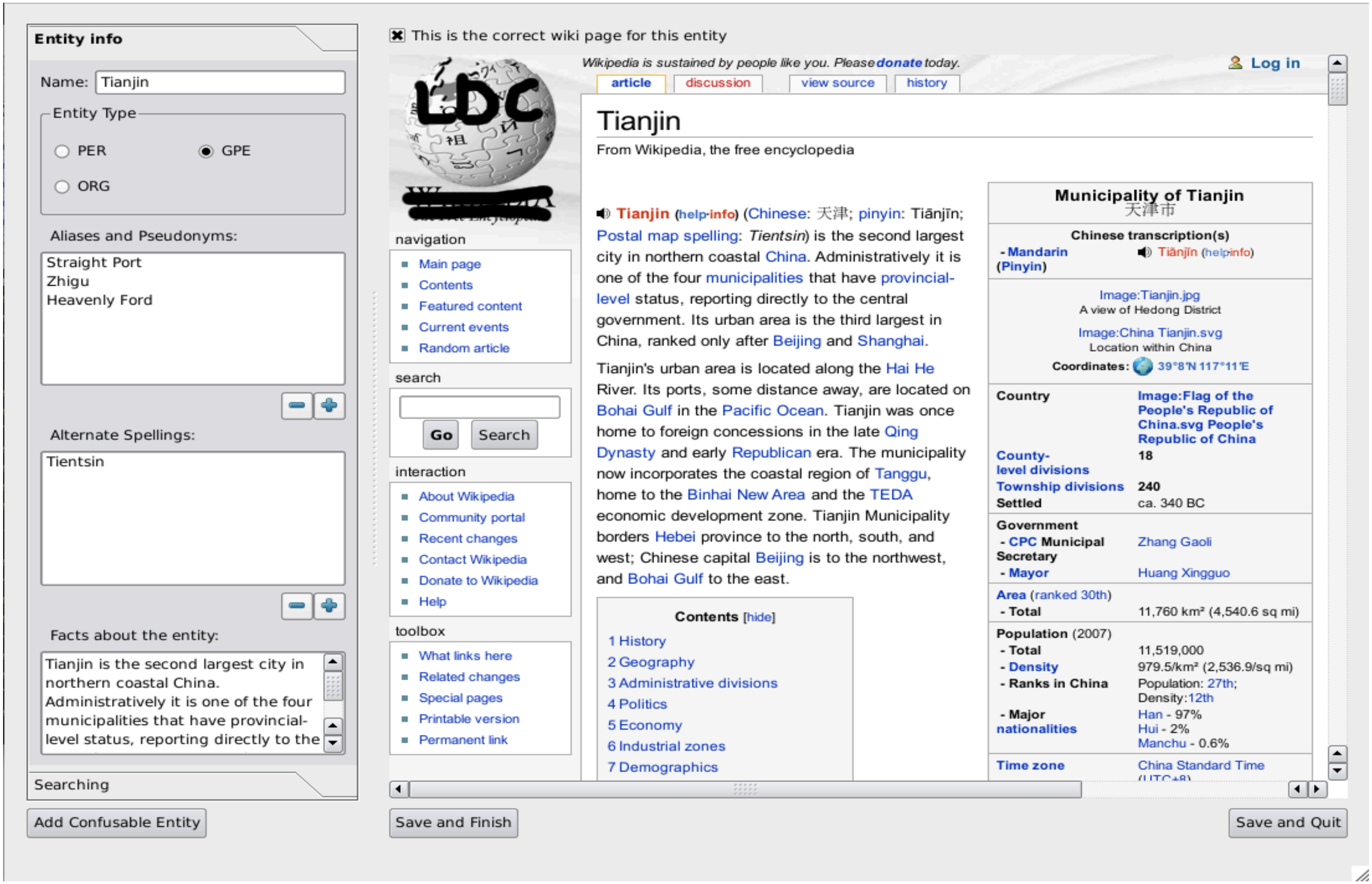
# Entity coverage in LDC-base vs. Knowledge Base



# Entity Linking Queries List

- Name mention-document pairs, GS links to Knowledge Base
  - desirable properties – variety, confusability, multiple name variants
- Developed via 2 stage process
  - Wikipedia Exploration Stage
    - Started with set of "seed" ACE profiles
    - Searched Wikipedia snapshot:
      - if matching entry found, link to LDC-base and add new facts/variants
      - if confusable, add new node to LDC-base and add facts/variants
  - Corpus Exploration Stage
    - Searched source data for name variants from Wikipedia exploration
    - Matched variants in document context to entity profiles
    - Created new entity profiles for variants not matching existing profiles

# Wikipedia Exploration Tool



# Corpus Exploration

Stop for now

## Assigned Variant: PCM

Done with annotation

**Entity Profile:** Communist Party of Madrid

**Type:** Organization

The federation of the Communist Party of Spain (PCE) in the Community of Madrid.

**Mention count:** 0

**Entity Profile:** Communist Party of Moldova

**Type:** Organization

Was one of the fourteen republic-level parties that formed the Communist Party of the Soviet Union until the fall of the Soviet Union in 1991. The PCM was the republic-level chapter of the CPSU in the Moldavian SSR from 1940 to 1991. During that time, except for the period of Axis occupation in 1941-1944, it was the sole legal political party in the republic. It was outlawed by the government in August 1991, just after Moldova declared independence.

**Mention count:** 0

**Entity Profile:** Malagasy Communist Party

**Type:** Organization

Malagasy Communist Party (French: Parti Communiste Malgache) was a communist party in Madagascar. PCM was formed in 1958 and its First Secretary was Rene Anselme Randrianja. It held its first party congress in Antanarivo March 18-March 20, 1960.

**Mention count:** 0

Add Confusable

**Falungong asks Canada FM's help to free 15 in China**

*OTTAWA, April 24, 2007 (AFP)*

Canadian relatives of 15 Falungong followers jailed in China asked on Tuesday Foreign Affairs Minister Peter MacKay to bring up their plight during an upcoming official visit to Beijing.

"The persecution of Falungong is a key policy of the Chinese regime among its many severe human rights violations," said Li Xun, president of the Falun Dafa Association of Canada.

"It is a serious issue that must be raised during any human rights talks with the regime."

China banned the spiritual group in 1999, accusing Falungong of spreading rumors in a bid to undermine "social stability" and Beijing's international relations, but the group is politically active in Canada.

The Canadian wing is expected to meet with the Foreign Affairs Department's China Desk on Wednesday to outline the plight of the 15, including the brother of a refugee who was himself freed from a Chinese prison with help from Canadian lawmakers, and the Beijing branch manager of Paris-based PCM Pumps.

The group accused Chinese authorities of "beating" and "brainwashing" their brethren, and sending them to forced labor camps.

Asylum seeker Yao Lian said PCM Pumps was "pressured to abandon their inquiries if they wished to continue doing business in China" after her husband Ma Jian was arrested at their Beijing offices.

"His arrest had a huge impact on their operations," she said, noting that French presidential candidate Segolene Royal wrote to Falun Dafa's Paris offices to "express concern" about Ma's fate, but failed to secure his release.

A Canadian government spokesman was not immediately available for comment.

But relations between Beijing and Ottawa have been strained recently over accusations that China is spying on Canadian corporations, the jailing of a Canadian imam in China, Canada's failure to deport a Chinese fugitive, talks between Ottawa and the Dalai Lama as well as stalled trade negotiations.

Last month, the wife of a Chinese diplomat defected to Canada and accused Beijing's embassy in Ottawa of inciting hatred against Falungong practitioners in Canada.

Previous | 1 of 7 | Next

# Entity Linking List

- Result from Wikipedia Exploration, Corpus Exploration, and Quality Control checks
  - 560 unique entities, 3904 name mention-document pairs (queries)
  - 15% PER, 70% ORG, 15% GPE
    - Original seed entities : 40% PER , 40% ORG, 20% GPE
  - 32.5 % KB, 67.5 % NIL
    - 33.4 % have 10/2008 Wikipedia entry with no infobox
    - 34.1% no 10/2008 Wikipedia entry
- query ID, name string, document ID
- Gold Standard version adds entity ID
  - Entity id = link to a unique entity node in KB or LDC-base (NIL)
  - used to evaluate performance on Entity Linking task

# Slot Filling

- Subset of Entity Linking task entities selected for slot filling task
  - Top goal : slot filling info in corpus but not in KB
  - Manual selection by lead annotator
    - Some KB, some NIL, variety of type
    - Entities with Wikipedia entries more newsworthy
      - 51% NIL entities with Wikipedia entries (33% in superset Entity Linking list)
    - KB entities with common info missing from infoboxes
      - Citibank : missing founded date, number of employees
- 53 entities, 32% PER, 58% ORG, 9% GPE

# Slot Filling Task Assessment

- Sequential assignment of all slots with pooled responses for an entity
- Stage 1: judge filler against doc vs. KB entry/Entity Profile
- Stage 2: for correct fillers, create equivalence classes
  - For entities already in KB, provide pre-existing equivalence class
- Stage 3: for correct fillers w/ proposed link to KB entry, judge the link
- Double- blind assessment, adjudication of disagreement
- NIL Link assessment  post-process
  - Searched KB for NIL links  in system output for Slot Filler KB linking task
  - 55/233 slot fillers had Wikipedia entries, 17/55 had infoboxes
  - Result: 17/233 assessed slot fillers found to be incorrectly NIL

# Conclusions

- Challenges
  - Name strings corresponding to multiple entities in documents
  - Specificity issue with GPE equivalence classes
  - Slot filling entities manual selection
- Suggestions
  - Automatic solution to multiple name strings in documents , or take char offsets for name strings
  - Build in stage to search corpus for slot fillers for subset of Entity Linking entities

# Entity Linking Metrics

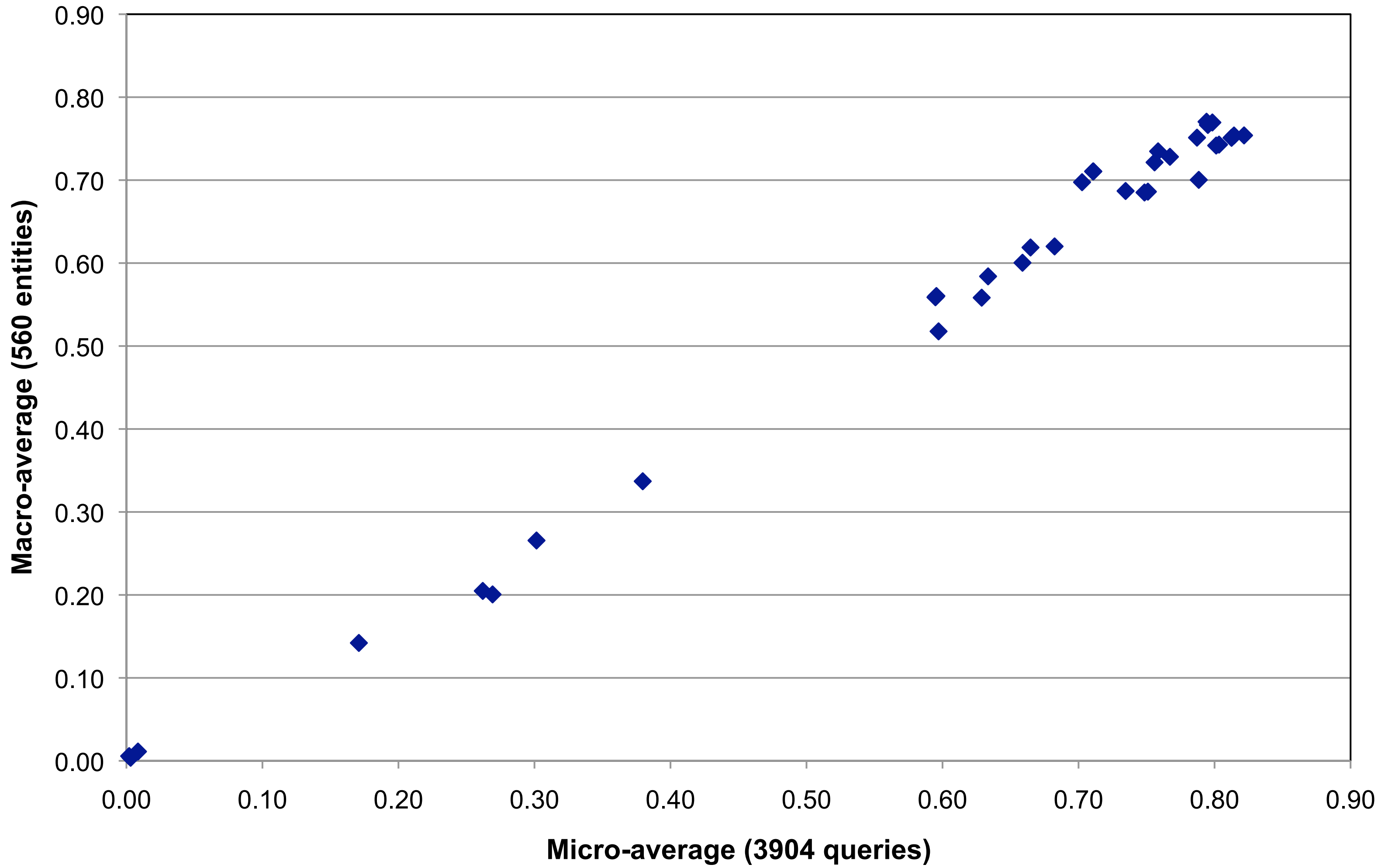
$$Accuracy_{micro} = \frac{NumCorrect}{NumQueries}$$

Estimate of performance for a random query. **Official Metric.**
3904 queries in total.

$$Accuracy_{macro} = \frac{\sum_{i}^{NumEntities} \frac{NumCorrect(E_i)}{NumQueries(E_i)}}{NumEntities}$$

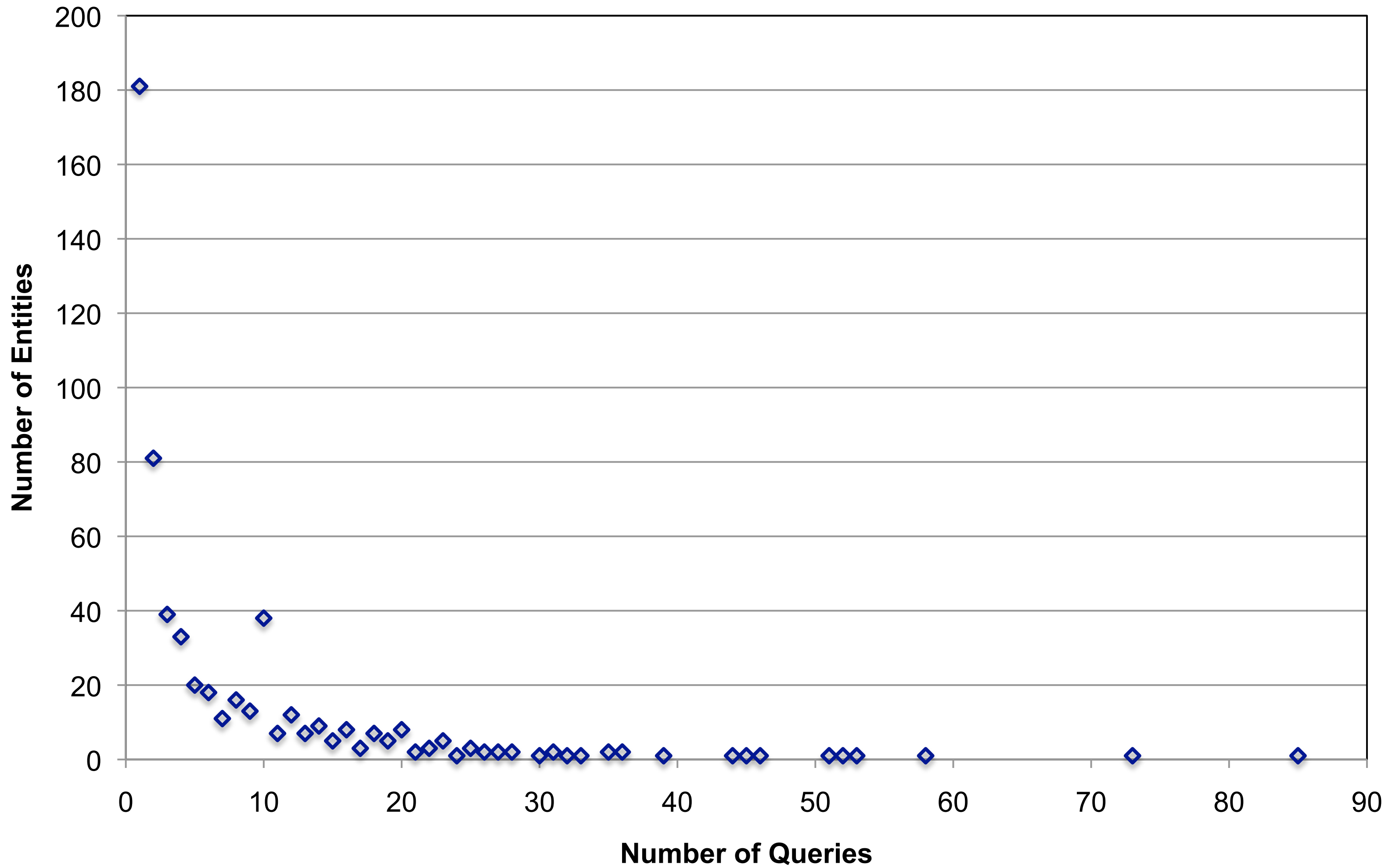
Estimate of performance for a random entity.
560 distinct entities.

# Micro vs. Macro




35 runs. Pearson correlation coefficient: 0.996

# Queries per Entity



# Top 5 Systems

| Team | All | in KB | NIL |
| --- | --- | --- | --- |
| Siel_093 | 0.8217 | 0.7654 | 0.8641 |
| QUANTA1 | 0.8033 | 0.7725 | 0.8264 |
| hltcoe1 | 0.7984 | 0.7063 | 0.8677 |
| Stanford_UBC2 | 0.7884 | 0.7588 | 0.8107 |
| NLPR_KBP1 | 0.7672 | 0.6925 | 0.8232 |
| | | | |
| 'NIL' Baseline | 0.5710 | 0.0000 | 1.0000 |

**Micro-averaged accuracy**

# Performance by Entity Type

| | All | in-KB | NIL |
|---|---|---|---|
| All | 0.8217 (3904) | 0.7654 (1675) | 0.8641 (2229) |
| | | | |
| PER | 0.8309 (627) | 0.8039 (255) | 0.8495 (372) |
| ORG | 0.8151 (2710) | 0.7305 (1013) | 0.8696 (1697) |
| GPE | 0.8480 (567) | 0.8280 (407) | 0.8812 (160) |

Performance for top-scoring run: Siel_093

# Hardest Queries

- Subsidiary organization
  - 3871 – Xinhua Finance Ltd .vs Xinhua Finance Media Ltd
- Typographical mistake / ambiguous acronym
  - 1213 – DCR for Democratic Republic of Congo
  - 3141 – MND (Taiwan Ministry of National Defense) referred to as NDM in text
- Metaphorical ‘names’
  - 1717/1718 Iron Lady (several strong female politicians)
- Unclear referent
  - 2599 – New Caledonia (country or soccer team)
- Mistakes in assessments
  - 3333,3334 – NYC Dept of Health, not US Dept of Health
  - 3335 – NY State Dept of of Health, not US Dept of Health

# Entity Linking Example

**EL1718 – Iron Lady**

**The furor also brought China's long-running domestic food safety problems to light, just as Beijing prepares to host hundreds of thousands of foreign visitors at the summer Olympics in August.**

**The seriousness with which the government took the issue was underscored by the appointment of its top problem solver, Vice Premier Wu Yi, to head a Cabinet-level panel overseeing the campaign.**

**Wu, a stern-looking, 69-year-old woman known as the "Iron Lady," shepherded China's difficult entry into the World Trade Organization, took over as health minister during the SARS epidemic and has been tasked with handling the vociferous U.S. complaints about China's exchange rate policy.**

**One month into the product safety campaign, Wu herself set out to randomly inspect shops and restaurants in the eastern province of Zhejiang. She had no itinerary and told no one in advance, making the driver stop at her whim.**

# Entity Linking Example

**EL3871 – Xinhua Finance**

**Chinese business news giant Xinhua Finance Media Ltd. is seeking to raise up to 371 million dollars through an initial public offering (IPO) on the Nasdaq stock market, according to a US regulatory filing.**

**...**

**"These outlets reach an estimated 210 million potential television viewers, a potential listening audience of 33 million people, and the readers of leading magazines and newspapers," Xinhua Finance Media said.**

**...**

**Describing itself as "a leading diversified media company in China," Xinhua Finance said it would use 50 million dollars from its US share listing to repay debts and "an undetermined amount" for future acquisitions.**

**The firm, which is based in the Cayman Islands, said it would be 36.7 percent owned by parent Xinhua Finance Ltd., 8.0 percent by Patriarch Partners Media Holdings LLC., and 5.8 percent owned by chief executive Fredy Bush, among other shareholders.**

## Sample SF Target

- **SF25: Convocation of Anglicans in North America**
  - **docid: LTW_ENG_20070506.0050.LDC2009T13**
  - **enttype: ORG**
  - **nodeid: NIL0031**

| Slot | Correct Values in Pools |
|---|---|
| org:alternate_names | CANA |
| org:founded | 2005 |
| org:founded_by | Peter Akinola |
| org:headquarters | Nigeria |
| org:member_of | Anglican Church, Nigerian Anglican Church |
| org:number_of_employees/members | 100,000 |
| org:parents | diocese of the Church of Nigeria, Nigerian Anglican Church |
| org:political/religious_affiliation | Anglican, Anglican Communion, Episcopal, Episcopal church, Christianity |
| org:top_members/employees | Peter Akinola, Bishop Martyn Minns, Kelly Oliver |
| org:website | www.canaconvocation.org |

# Convocation of Anglicans in North America

- **founded_by**
  - **Akinola, AMIA Bishop Chuck Murphy, Bishop Martyn Minns, Episcopal, Helmandollar, Jim Robb, Martyn Minns, Minns, Peter Akinola, Robinson, Stephen**
- **shareholders**
  - **Anglican Church, Bishop Martyn Minns, CANA, Episcopal Church, Martyn Minns, Peter Akinola**
- **headquarters**
  - **America, Nigeria, Quincy, Woodbridge**

# Slot Filling Scoring

- Responses were marked as one of Correct, Inexact, Redundant, or Wrong
- Responses had to be justified from a single supporting document
  - Unsupported responses were marked wrong

- 53 target entities (17 PER, 31 ORG, 5 GPE)
  - 255 single-value slots – 39 (15%) had correct values in the pooled responses
  - 499 list slots – 129 (26%) had correct values
  - Thus predicting NIL (no response) is correct ~80% of the time
  - 48/53 entities had at least one learnable attribute

# Easy / Hard Slots

| Slot | Filled Entities | Correct Responses | Submitted Responses |
|---|---|---|---|
| per:title | 16/17 | 86 | 409 |
| per:employee_of | 10/17 | 38 | 429 |
| per:origin | 9/16 | 16 | 117 |
| per:member_of | 9/17 | 41 | 424 |
| org:top_members/ employees | 24/31 | 258 | 1463 |
| org:alternate_names | 23/31 | 87 | 710 |
| org:headquarters | 11/21 | 17 | 131 |

- **No values for:**
  - **PER: other_family, parents, spouse**
  - **ORG: shareholders**
  - **GPE: capital, political_parties, population**

# Slot Filling Metrics

$$Score_{\text{single}} = \frac{NumCorrect}{NumSingleSlots}$$

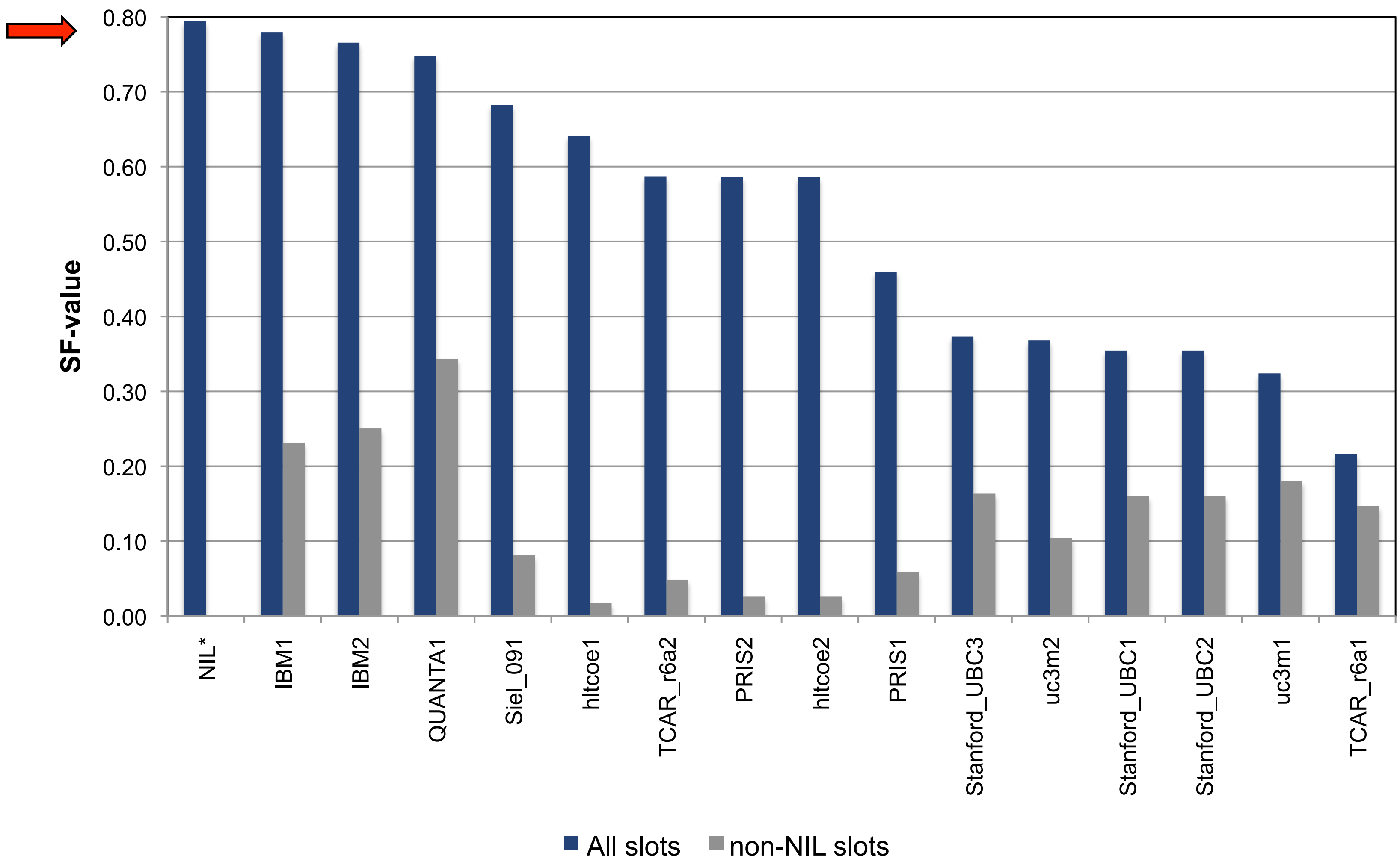
$$ListSlotValue = \frac{5 \times IP \times IR}{4 \times IR + IP}$$

$F_\beta$=2 to weight precision over recall.
IP = Instance precision.
IR = Instance recall.

$$Score_{\text{list}} = \frac{\sum ListSlotValue}{NumListSlots}$$

$$SF_{\text{value}} = \tfrac{1}{2}\left(Score_{\text{single}} + Score_{list}\right)$$

# SF Results



0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
SF-value
NIL*
IBM1
IBM2
QUANTA1
Siel_091
hltcoe1
TCAR_r6a2
PRIS2
hltcoe2
PRIS1
Stanford_UBC3
uc3m2
Stanford_UBC1
Stanford_UBC2
uc3m1
TCAR_r6a1
All slots
non-NIL slots

# Lessons from Slot Filling

- GPEs have few learnable attributes in news
  - latitude, longitude, elevation not commonly reported
  - population is, but usually available in KB/Wikipedia
- Difficult to estimate how much information is available (and novel) for a candidate target entity
  - Manual search needed both to facilitate target selection and enrich pools
- Balance scoring between slots with discoverable vs. NIL values
- End-to-end assessment of ‘KB improvement’ is difficult. Component evaluation for KBP is worth considering.
  - Can a passage support a given slot for a given entity? (The IR4QA problem)
  - Is a particular slot fill justified from a passage? (An RTE task)
  - Is this slot fill redundant with another value?

# Evaluation Issues

- **Imperfect KB**
  - **Wikipedia focuses on presentation, not representation**
    - **irrelevant slots (colors, image sizes), values are not normalized (e.g., dates)**
  - **Many non-entities**
- **Use of external resources**
- **Generic entities (vs. thousands of classes)**
  - **Slot names were inconsistent (birthdate, date-of-birth)**
- **Response granularity**
  - **USA, Hawaii, Honolulu – which should be considered correct birthplaces for President Obama?**
- **Dealing with time**
  - **Key USA leadership: G. Washington or B. Obama**
- **Query Difficulty (and high NIL percentage)**
- **Assessing KB Growth**
  - **Difficult to directly measure benefit from adding to KB**

# Conclusion

- Pilot evaluation for adding information to a reference knowledge base
- 2 initial tasks
  - Linking name mentions to KB entries
  - Augmenting profiles for target entities
- KBP 2010
  - Refine and extend evaluation
  - Ralph Grishman and Heng Ji have volunteered to serve as the track coordinators
  - Please come to the planning meeting!