# Overview of the TAC2011 Knowledge Base Population (KBP) Track

Heng Ji, Ralph Grishman and Hoa Trang Dang

November 15, 2011

# Goal of KBP

- General Goal
  - Promote research in discovering facts about entities to create and expand a knowledge source automatically
- What's New in 2011
  - Support multi-lingual information fusion – a new Cross-lingual Entity Linking task
  - Capture temporal information – a new Temporal Slot Filling task
  - Added clustering of entity mentions without Knowledge Base entries into the Entity Linking task, and developed a new scoring metric incorporating NIL clustering
  - Made systematic corrections to the slot filling guidelines and data annotation
  - Defined a new task, Cross-lingual Slot Filling, and prepared its annotation guideline

# KBP Setup

Source Collection

*Create/Expand*

Reference KB

CNN U.S.

SEARCH

POWERED BY Google

Justice | Entertainment | Tech | Health | Living | Travel | Opinion | iReport | Money | Sports

## Steve Jobs, Apple founder, dies

October 05, 2011 | By Brandon Griggs, CNN

Share | Twitter | Email

Recommend | 153,356 people recommend this.

Steve Jobs, the visionary in the black turtleneck who co-founded Apple in a Silicon Valley garage, built it into the world's leading tech company and led a mobile-computing revolution with wildly popular devices such as the iPhone, died Wednesday. He was 56.

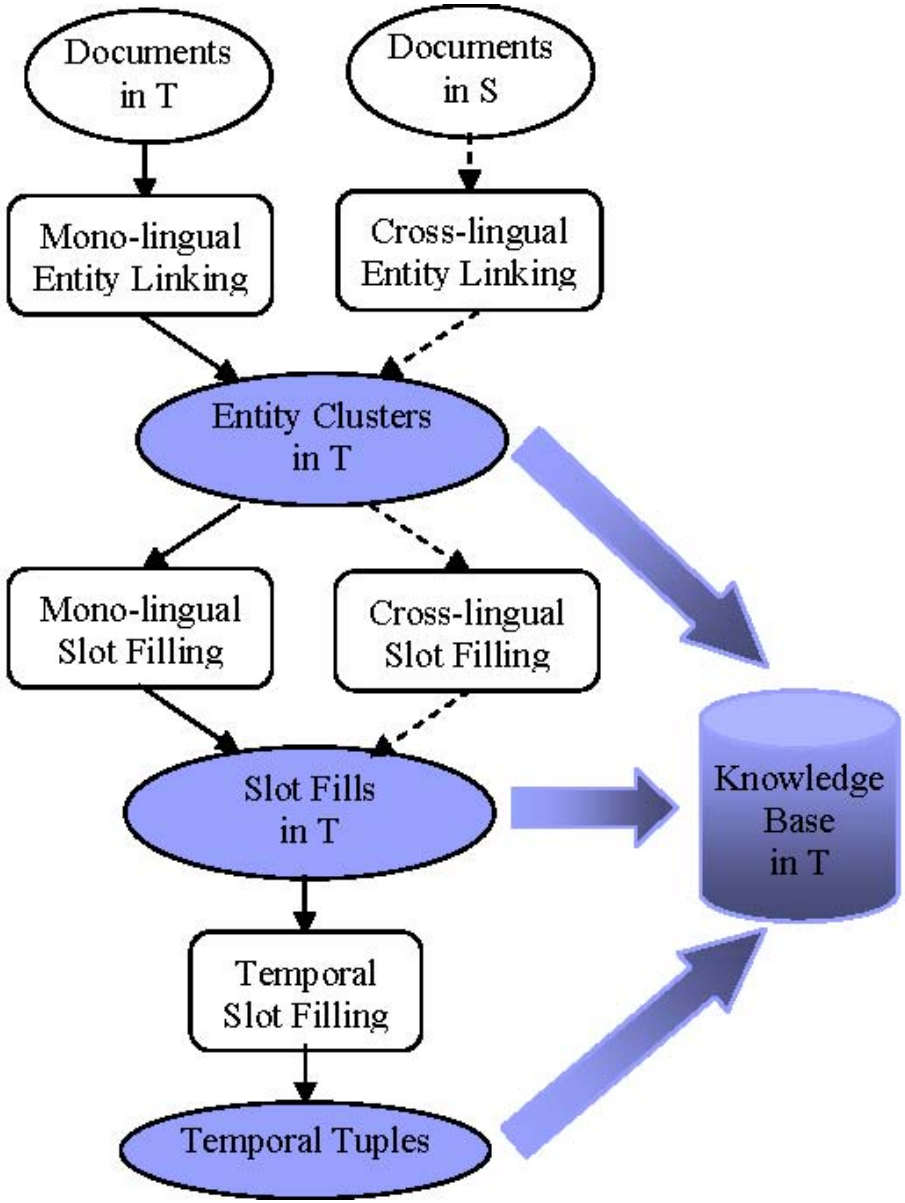
The hard-driving executive pioneered the concept of the personal computer and of navigating them by clicking onscreen images with a mouse. In more recent years, he introduced the iPod portable music player, the iPhone and the iPad tablet -- all of which changed how we consume content in the digital age.

**Steve Jobs**

Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

| | |
|---|---|
| **Born** | Steven Paul Jobs<br>February 24, 1955[1][2]<br>San Francisco, California, U.S.[1][2] |
| **Died** | October 5, 2011 (aged 56)[2]<br>Palo Alto, California, U.S. |
| **Nationality** | American |
| **Occupation** | Co-founder, Chairman and CEO, Apple Inc., CEO, Pixar, Co-founder and CEO, NeXT Inc. |
| **Years active** | 1974–2011 |

# Overview of KBP Tasks



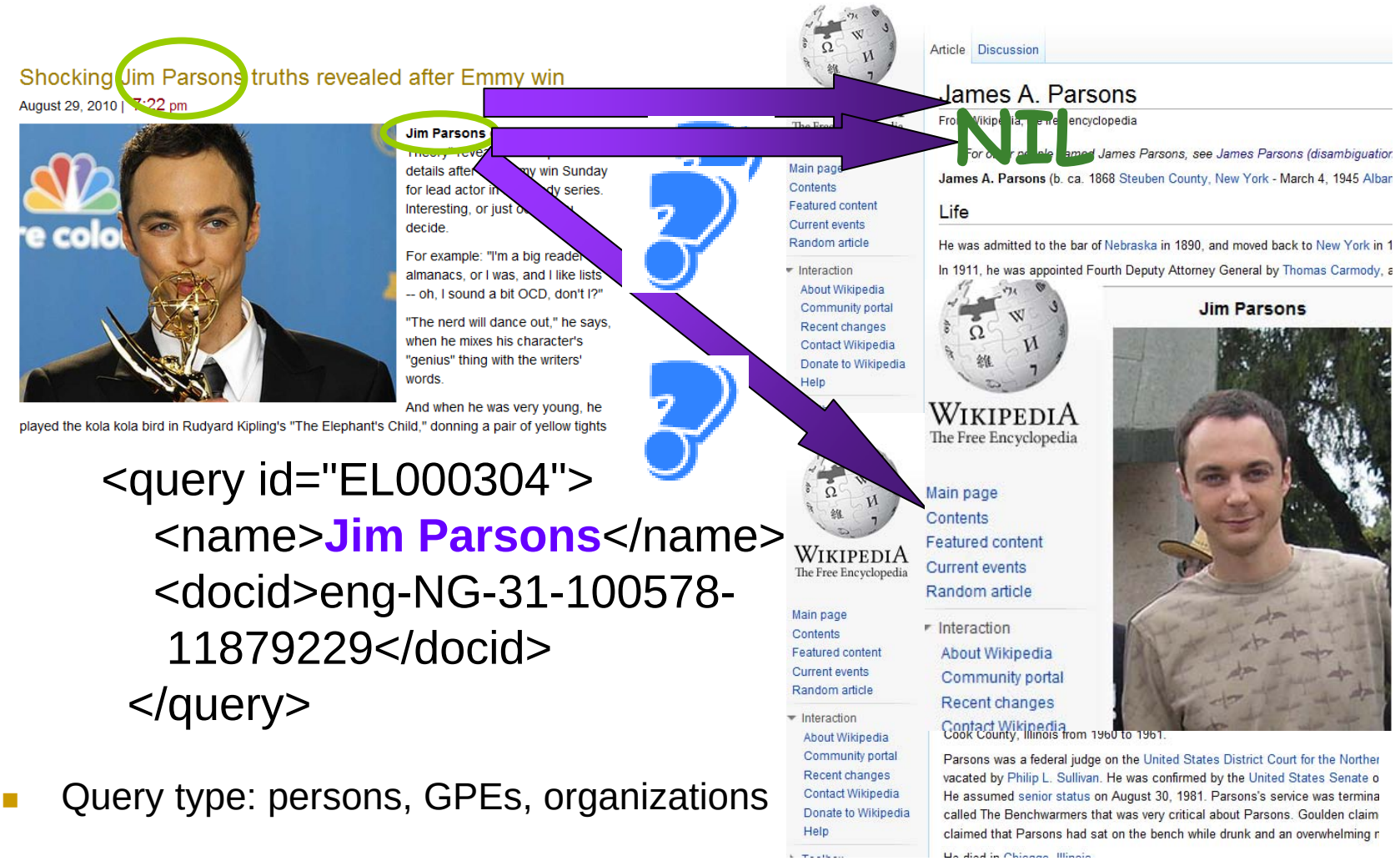
# KBP2011 Participants

- 65 teams registered for KBP 2011 (not including the RTE-KBP Pilot task), 35 teams submitted results
- Each team can submit up to 3 submissions

| Task / Participants/Year | | Entity Linking | | | Slot Filling | | | |
|---|---|---|---|---|---|---|---|---|
| | | Mono-lingual | | Cross-lingual | Regular | Surprise | Temporal | |
| | | Regular | Optional | | | | Full | Diagnostic |
| #Teams | 2009 | 13 | - | - | 8 | - | - | - |
| | 2010 | 16 | 7 | - | 15 | 5 | - | - |
| | 2011 | **22** | **8** | **11** | 14 | - | **5** | **4** |
| #Submissions | 2009 | 35 | - | - | 16 | - | - | - |
| | 2010 | 46 | 20 | - | 31 | 6 | - | - |
| | 2011 | **53** | 15 | **27** | 31 | - | **11** | **7** |

# I: Mono-lingual Entity Linking

# Entity Linking: Create Wiki Entry?




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

- Query type: persons, GPEs, organizations

# Entity Linking Scoring Metric: B-cubed+

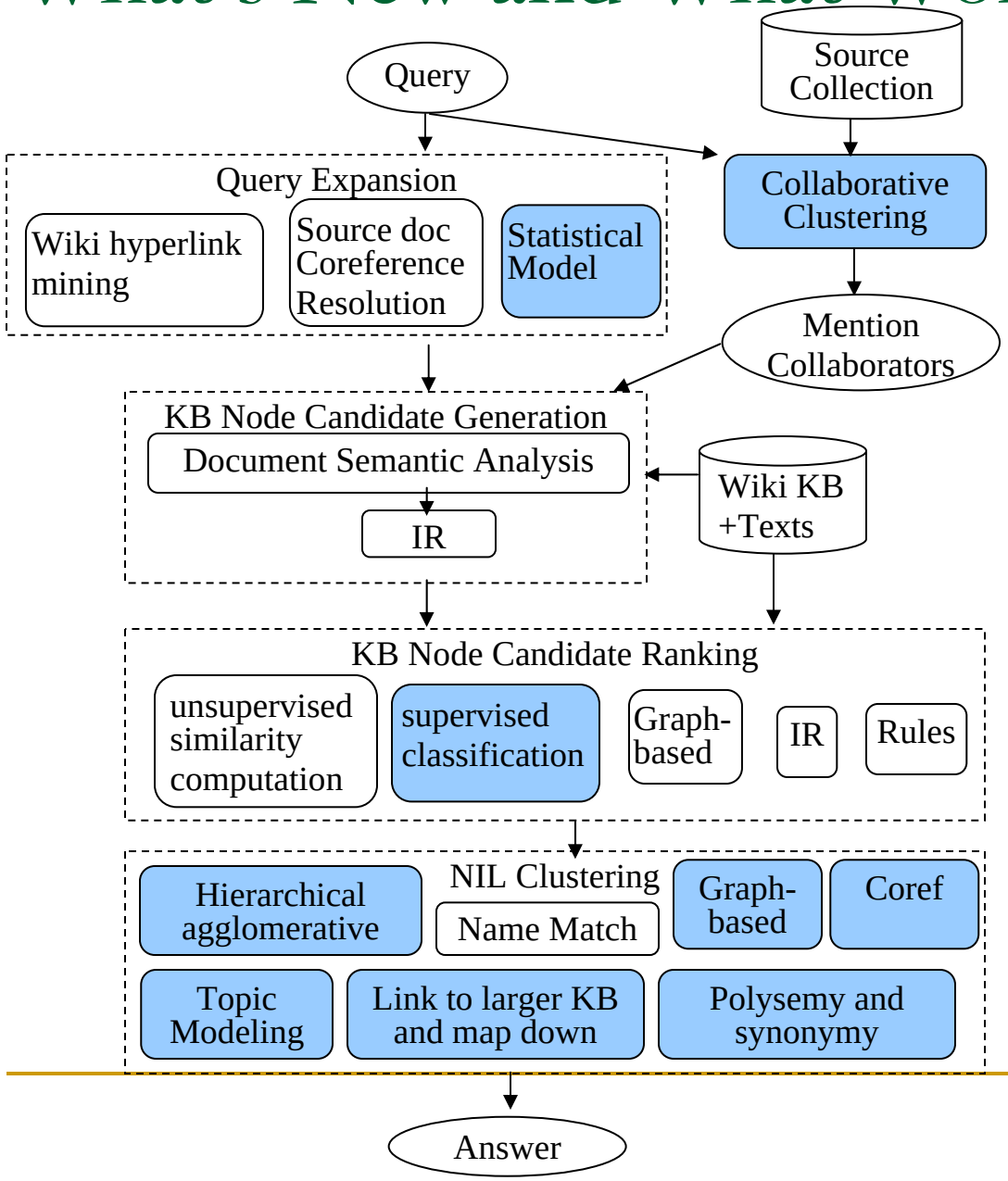
- $L(e)$ and $C(e)$: the category and the cluster of an item $e$
- $SI(e)$ and $GI(e)$ : the system and gold-standard KB identifier for an item $e$
- The correctness of the relation between $e$ and $e'$ in the distribution:

$$G(e,e') = \begin{cases} 1 iff\ L(e) = L(e') \wedge C(e) = C(e') \wedge GI(e) = SI(e) = GI(e') = SI(e') \\ 0\, otherwise \end{cases}$$

$$\Pr ecision\ B-Cubed+ = Avg_e[Avg_{e'.C(e)=C(e')}[G(e,e')]]$$

$$\mathrm{Re} call\ B-Cubed+ = Avg_e[Avg_{e'.L(e)=L(e')}[G(e,e')]]$$

# What's New and What Works

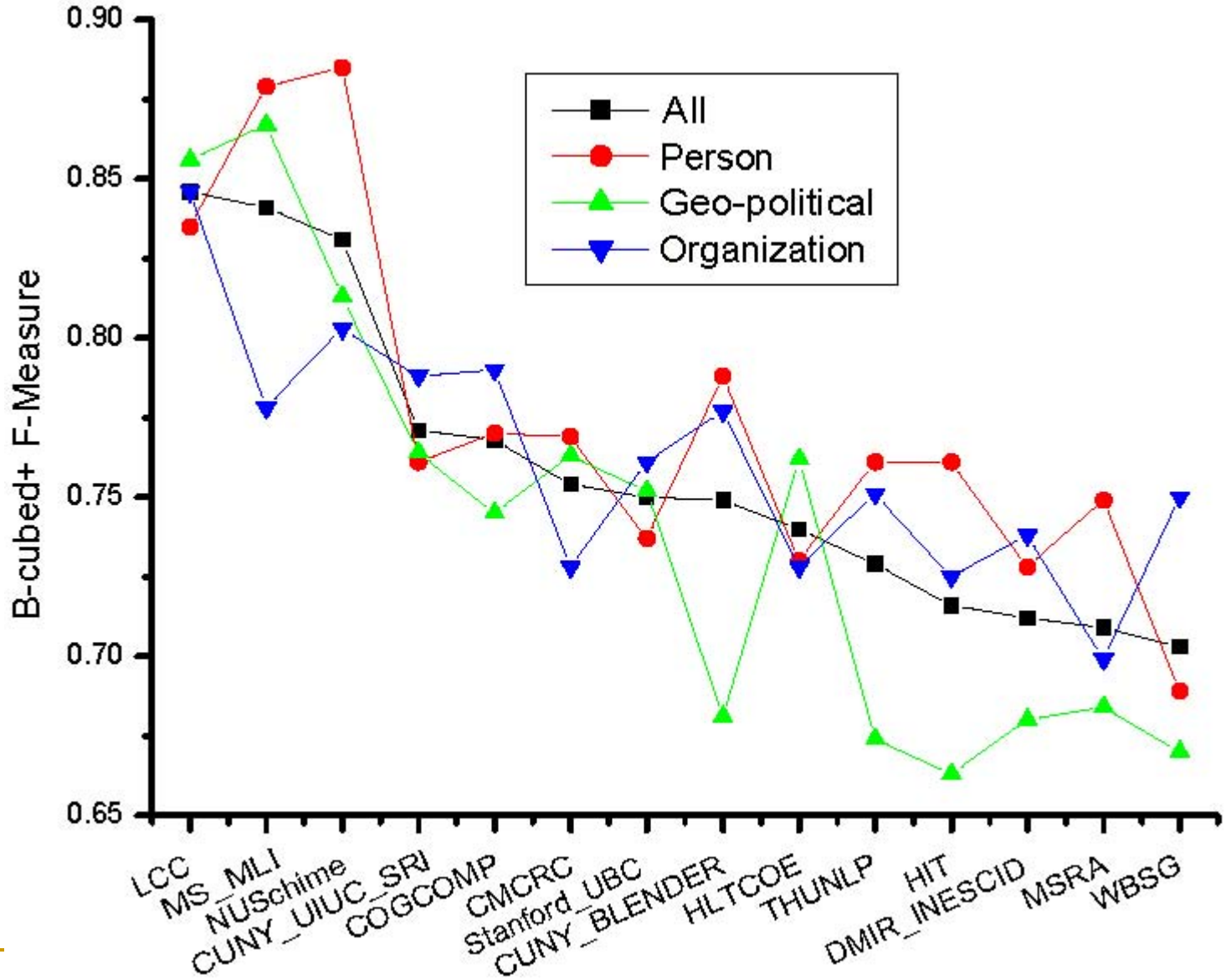



- Statistical Name Variant Expansion (NUSchime)
  - "CCP" vs. "Communist Party of China"
  - "MINDEF" vs. "Ministry of Defence"
- New Ranking Algorithms
  - e.g. ListNet (CUNY), Random Forests (THUNLP,DMIR_INESCID)
- Query Classification
  - DMIR_INESCID, CUNY, MSRA
- Go Beyond Single Query and Single KB Entry
  - Wikification (UIUC), Collaborative ranking (CUNY), Link all entities and inference (MS_MLI, CMCRC)

# Typical Ranking Features

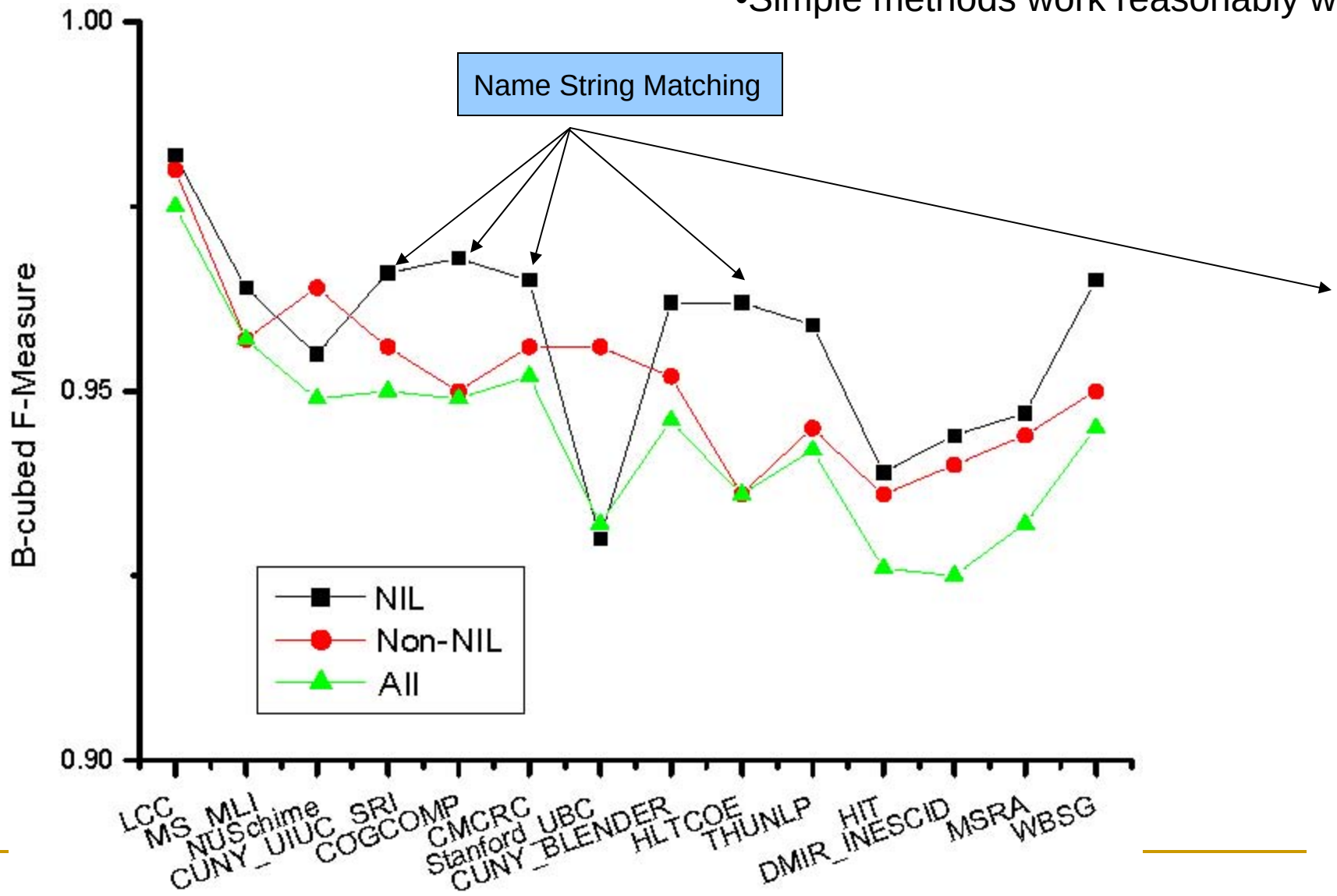
| Feature Category | | Feature Description |
|---|---|---|
| Name | Spelling match | Exact string match, acronym match, alias match, string matching… |
| | KB link mining | Name pairs mined from KB text redirect and disambiguation pages |
| | Name Gazetteer | Organization and geo-political entity abbreviation gazetteers |
| Document surface | Lexical | Words in KB facts, KB text, query name, query text. |
| | | Tf.idf of words and ngrams |
| | Position | Query name appears early in KB text |
| | Genre | Genre of the query text (newswire, blog, …) |
| | Local Context | Lexical and part-of-speech tags of context words |
| Entity Context | Type | Query entity type, subtype |
| | Relation | Entities co-occurred, attributes/relations/events with the query |
| | Coreference | Coreference links between the source document and the KB text |
| Profile | | Slot fills of the query, KB attributes |
| Concept | | Ontology extracted from KB text |
| Topic | | Topics (identity and lexical similarity) for the query text and KB text |
| KB Link Mining | | Attributes extracted from hyperlink graphs of the KB text |
| Popularity | Web | Top KB text ranked by search engine and its length |
| | Frequency | Frequency in KB texts |

# Top MLEL System Performance (Regular Task)

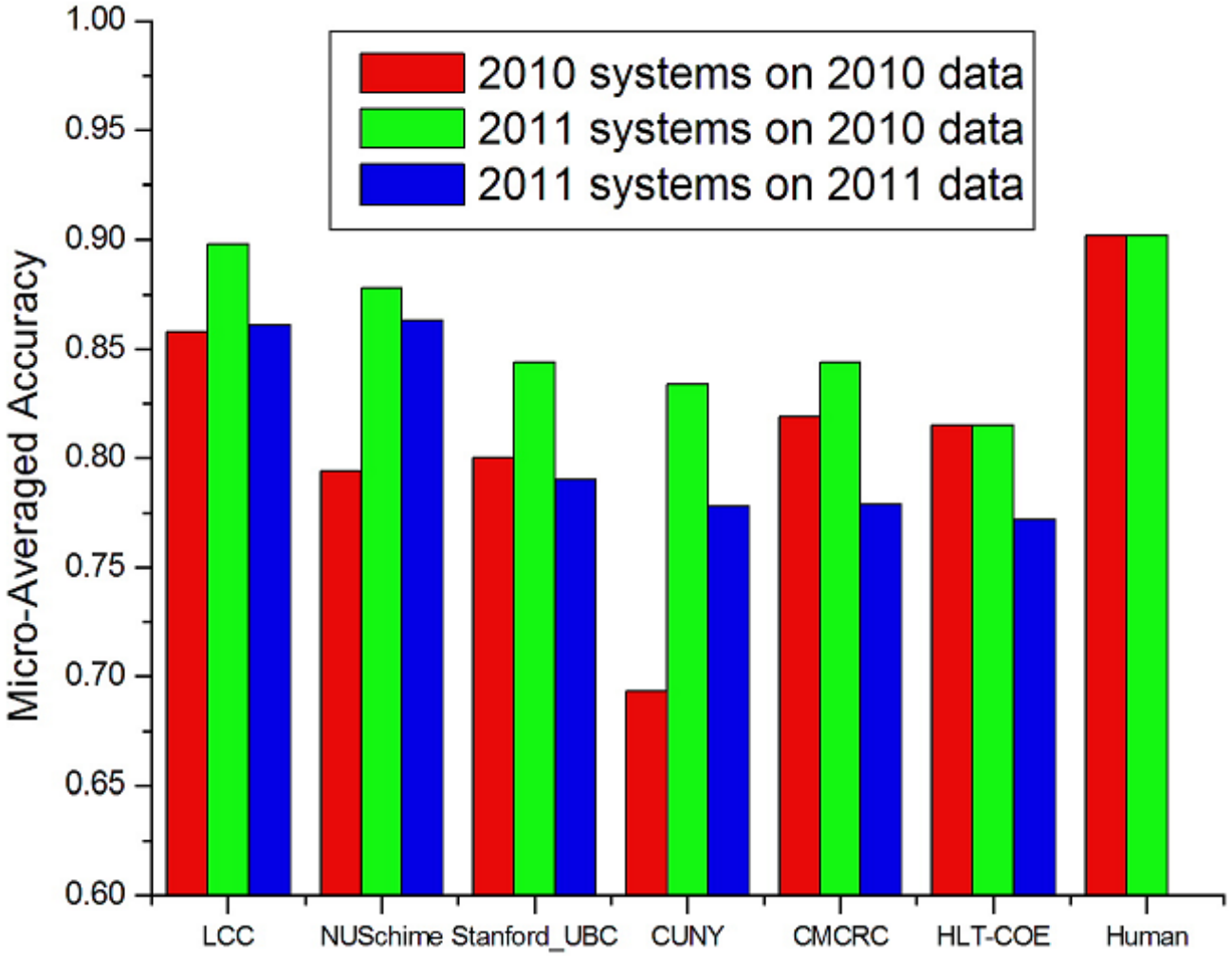


# MLEL NIL Clustering Performance

•Simple methods work reasonably well



# Progress of Top MLEL Systems




ambiguity = % of name strings which refer to more than one cluster

2010: **5.7%** vs.
2011: **12.1%**

# II: Cross-lingual Entity Linking

# Cross-lingual Entity Linking

<query id="SF114">
<name>李安</name>
<docid>XIN20030616.0130.0053</docid>
</query>


李安 – 简介 纠错 | 编辑本段

**Parent: Li Sheng**

李安，台湾著名导演，祖籍江西省九江市德安县，生于台湾屏东县，父亲李升。李安高中原就读台南二中，后转学考进了台南第一志愿——台南一中。对于读书，李安一点兴趣都没有，心里只想着当导演。大学考试落榜两次，后来准备专科考试，进了国立台湾艺专（今国立台湾艺术大学）影剧科，从此改变了李安的一生。

**Birth-place: Taiwan Pindong City**

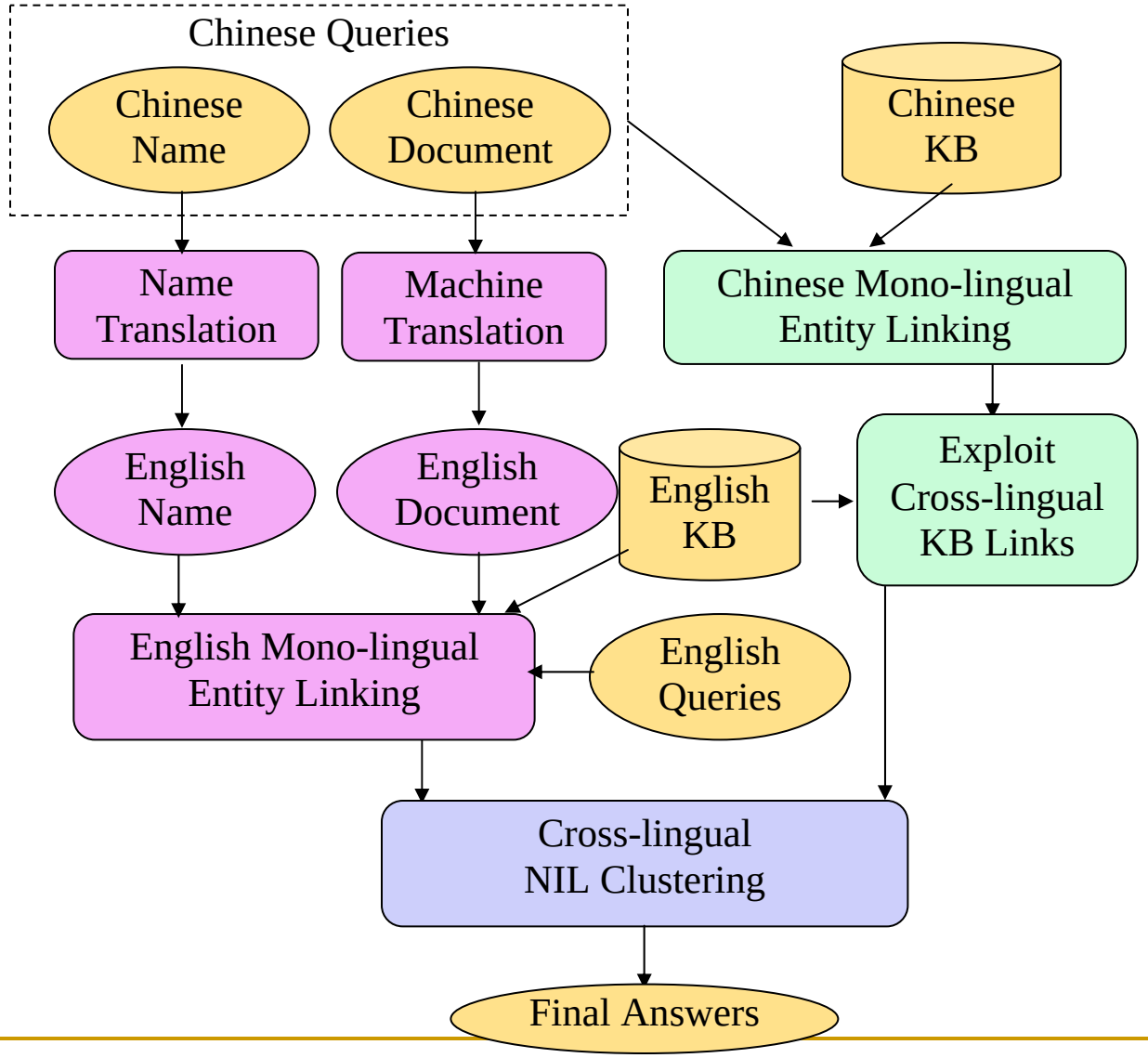
**Residence: Hua Lian**

李安曾言，住在花莲的八年，乃其北上就读艺专前最快乐的一段学习岁月。十岁之前的李安在花莲念了两所小学，接受的是美式开放教育，来到台南，又念了两所小学，面对语言习惯不同国语－台语，头一次经验到文化冲击。

**Attended-School: NYU**

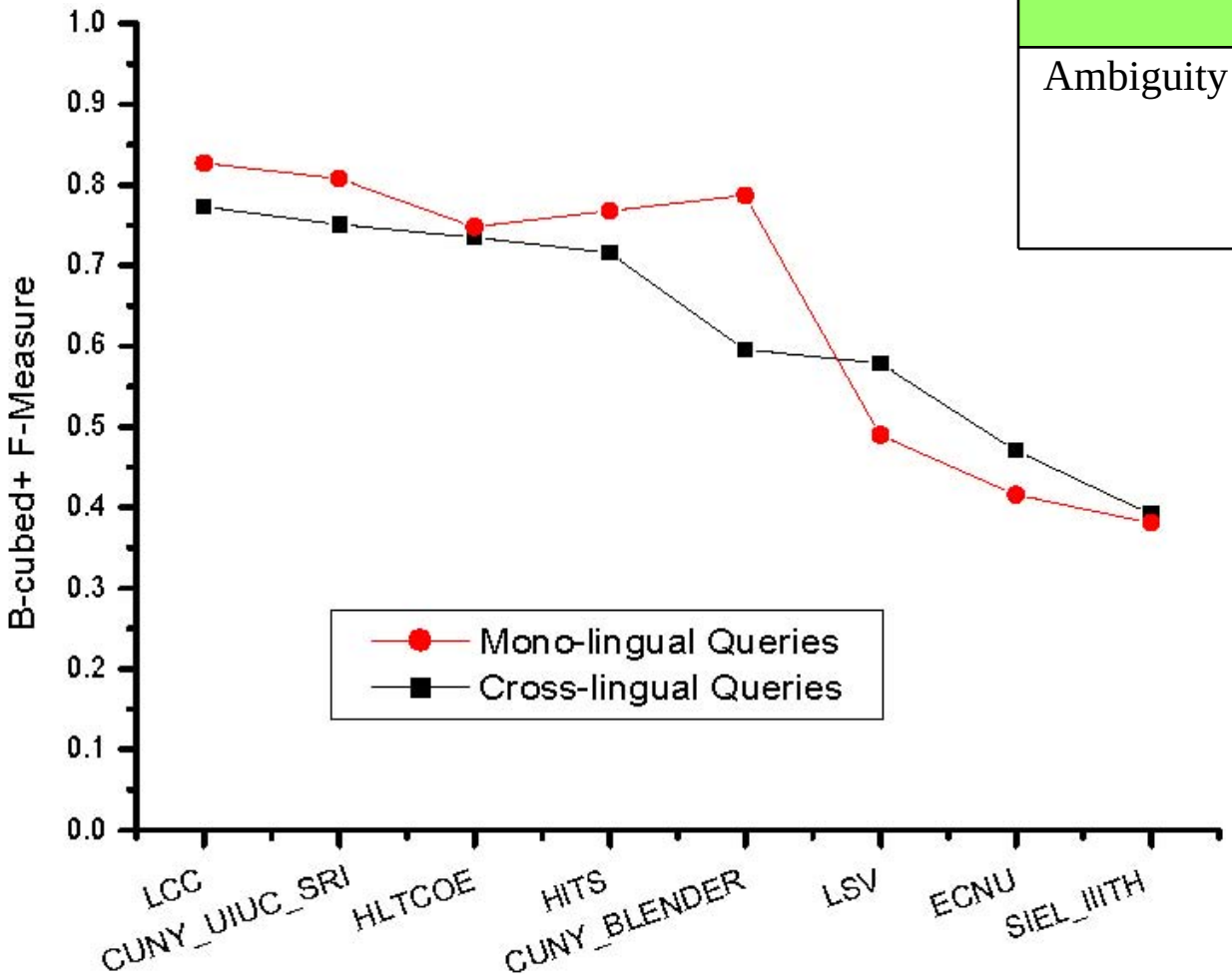
李安于1979年赴美就读伊利诺大学香槟分校戏剧系取得学士学位，后于1981年至纽约大学就读电影制作研究所，取得硕士学位。李安的妻子林惠嘉是伊利诺大学香槟分校生物学博士，现任纽约医学院病理学研究员。

# General CLEL System Architecture

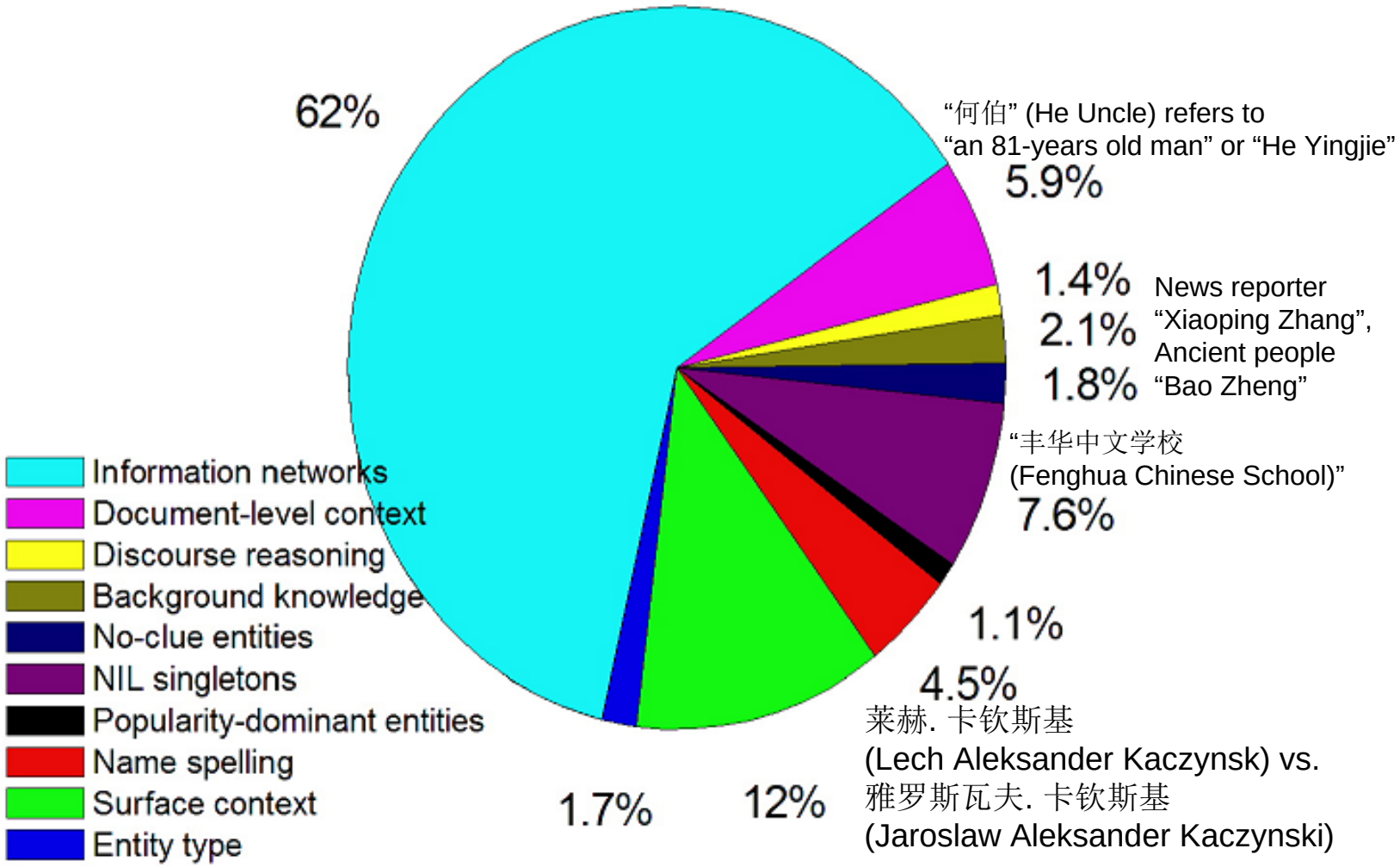


# From Mono-lingual to Cross-lingual

| Difficulty | Task | All | NIL | Non-NIL |
|---|---|---|---|---|
| Ambiguity | Mono-lingual | 12.9 % | 5.7 % | 9.3% |
| | Cross-lingual | 20.9 % | 14.0 % | 28.6 % |



# CLEL Knowledge Categorization



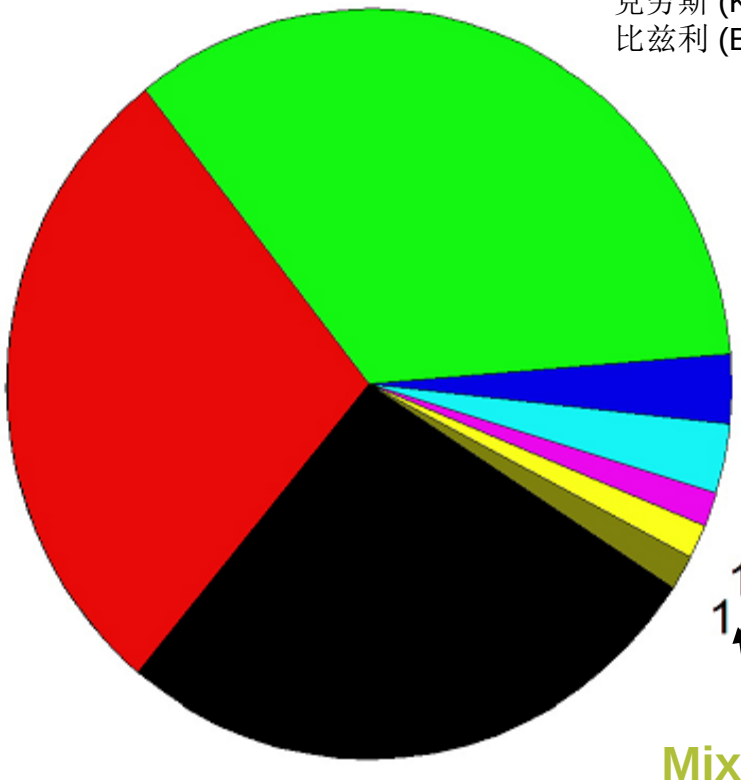
# Person Name Translation Challenges

**Name Transliteration + Global Validation**: 34%

克劳斯 (Klaus), 莫科(Moco)
比兹利 (Beazley), 皮耶 (Pierre)…

**Name Pair Mining and Matching (common foreign names)** 28%

伊莎贝拉 (Isabella), 斯诺(Snow),
林肯(Lincoln), 亚当斯(Adams)…



**Pronunciation vs. Meaning confusion** 3%

拉索 (Lasso vs. Cable)
何伯 (He Uncle)

**Entity type confusion** 3%

魏玛 (Weimar vs. Weima)

**Chinese Name vs. Foreign Name confusion** 1.5%

洪森 (Hun Sen vs. Hussein)

**Origin confusion** 1.5%

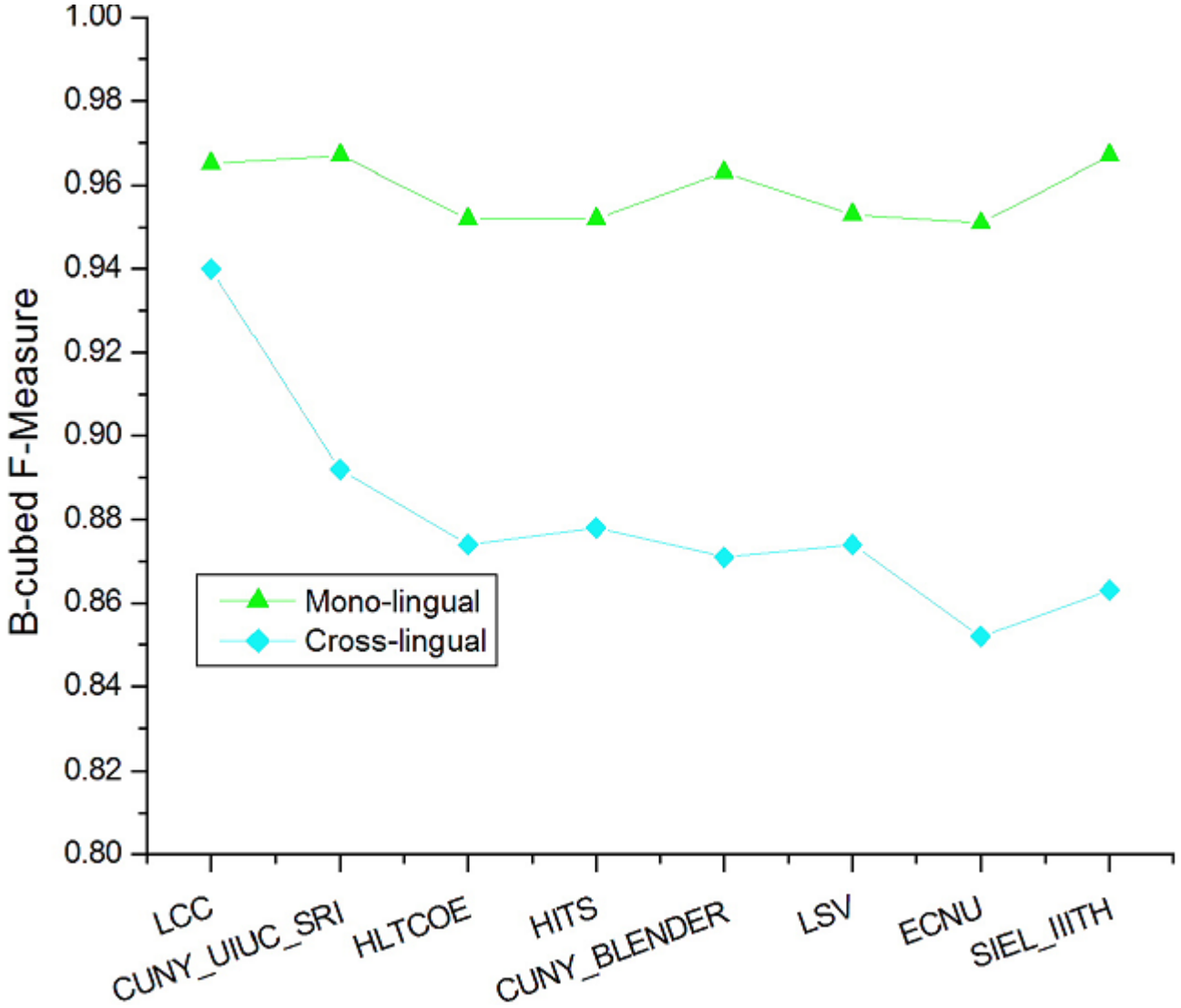
**Mixture of Chinese Name vs. English Name** 1.5%

王菲 (Faye Wong)

**Chinese Names (Pinyin)** 27%

王其江 (Wang Qijiang), 吴鹏(Wu Peng), …

# CLEL NIL Clustering Performance



# Cross-lingual NIL Clustering

- One-to-Many Clustering
  - Li Na, Wallace, …
- Topic Modeling Errors
  - The same name (莫里西/Molish), the same topic (life length/death analysis), different entities
- Require temporal employment tracking
  - 众议院情报委员会主席高斯 (Gauss, the chairman of the Intelligence Committee) =美国中央情报局局长高斯 (The U.S. CIA director Gauss)

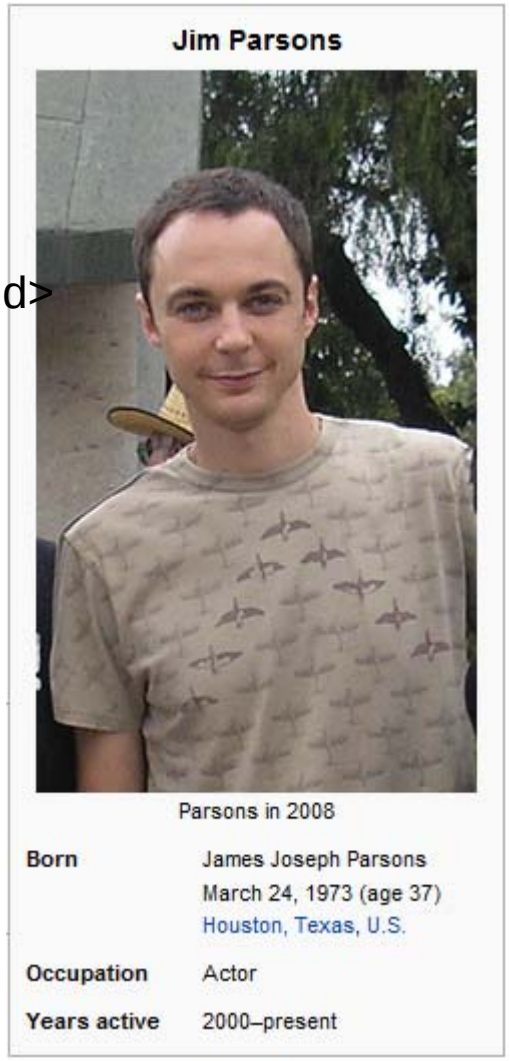
# III: Regular Slot Filling

# Regular Slot Filling

```
<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 [illegible] and Dance, won the Emmy on Sunday for Lead Actor in a Comedy Series for his work on The Big Bang Theory.

**School Attended: University of Houston**



# Attribute Distribution in Regular Slot Filling

| ORG slot | values | PER slot | values | PER slot | values |
|---|---|---|---|---|---|
| **top_members, employees** | **118 (12%)** | **title** | **201 (21%)** | country_of_death | 1 (0%) |
| **alternate names** | **98 (10%)** | **employee_of** | **71 (7%)** | date_of_birth | 3 (0%) |
| **subsidiaries** | **32 (3%)** | **alternate_names** | **46 (4%)** | date_of_death | 4 (0%) |
| country of headquarters | 22 (2%) | **member_of** | **47 (4%)** | city_of_death | 1 (0%) |
| org:parents | 24 (2%) | countries_of_residence | 20 (2%) | city_of_birth | 6 (0%) |
| member_of | 11 (1%) | origin | 23 (2%) | country_of_birth | 3 (0%) |
| shareholders | 18 (1%) | charges | 15 (1%) | other_family | 6 (0%) |
| stateorprovince_of_headquarters | 17 (1%) | children | 17 (1%) | parents | 3 (0%) |
| city of headquarters | 19 (1%) | cities_of_residence | 17 (1%) | religion | 5 (0%) |
| website | 14 (1%) | age | 16 (1%) | siblings | 6 (0%) |
| political,religious_affiliation | 2 (0%) | schools_attended | 16 (1%) | spouse | 8 (0%) |
| dissolved | 1 (0%) | stateorprovinces_of_residence | 11 (1%) | stateorprovince_of_birth | 1 (0%) |
| members | 8 (0%) | | | cause_of_death | 3 (0%) |
| number_of_employees,members | 6 (0%) | | | | |
| founded | 6 (0%) | | | | |
| founded_by | 7 (0%) | | | | |

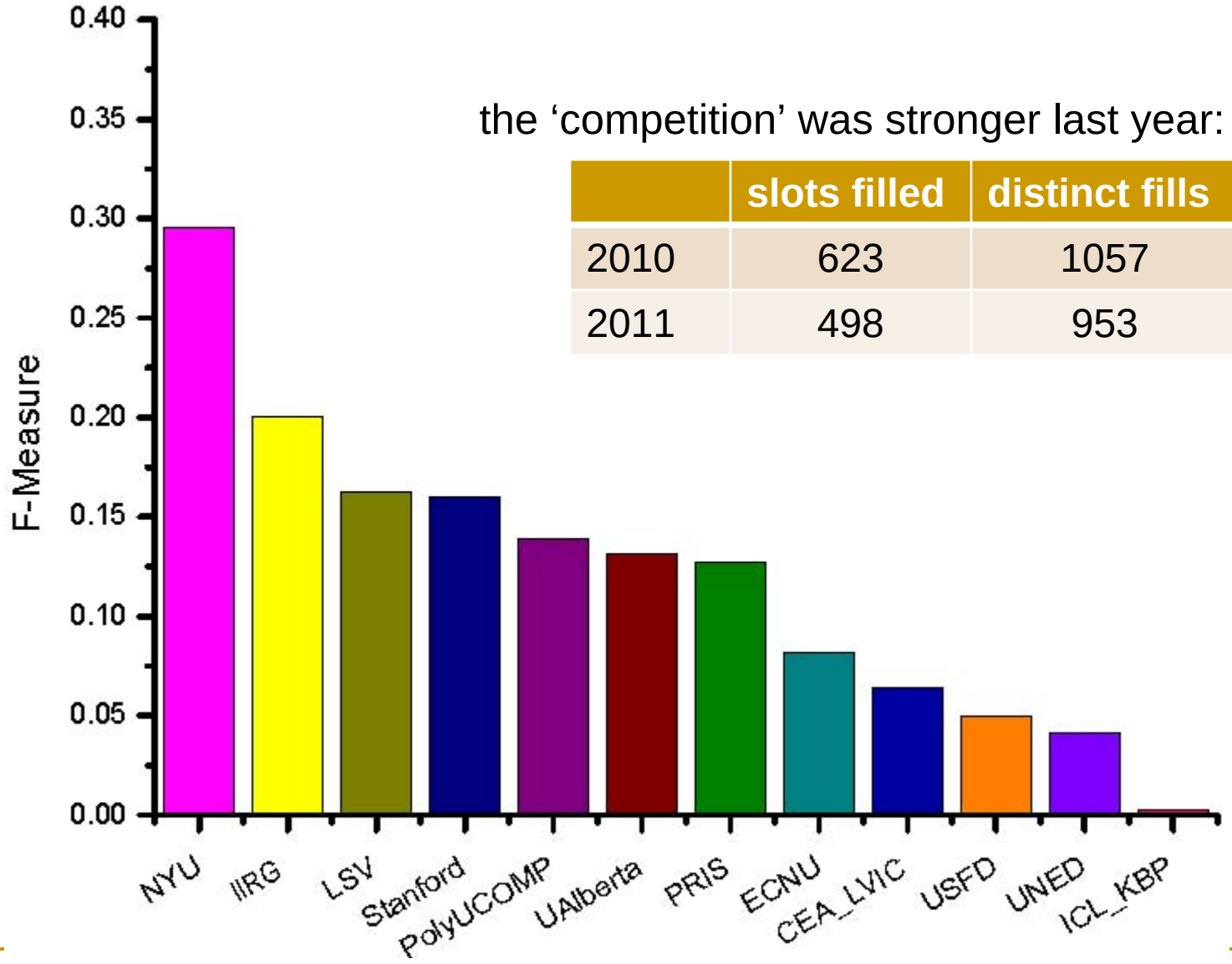
# Regular Slot Filling Scoring Metric

- 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

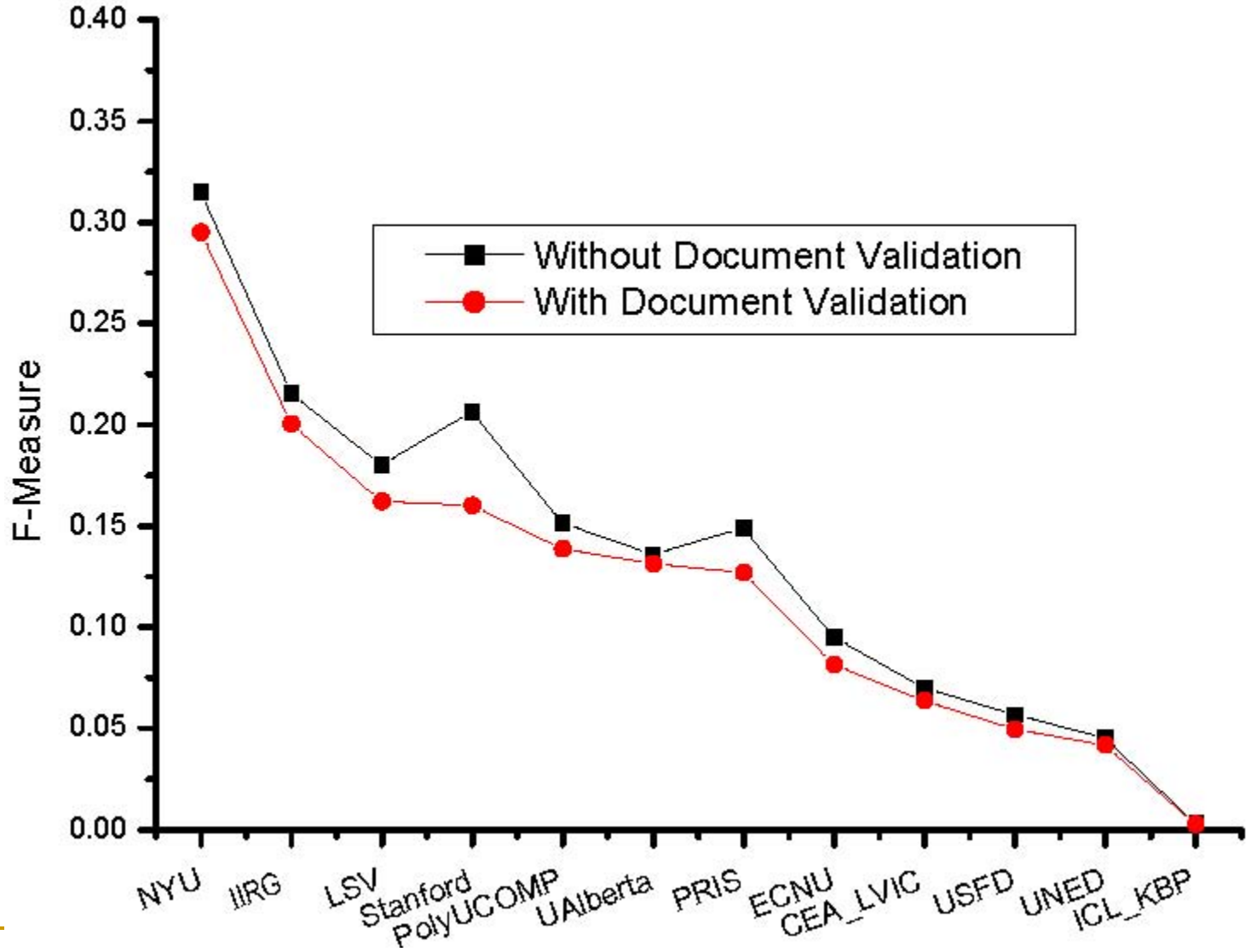
# Regular Slot Filling Systems



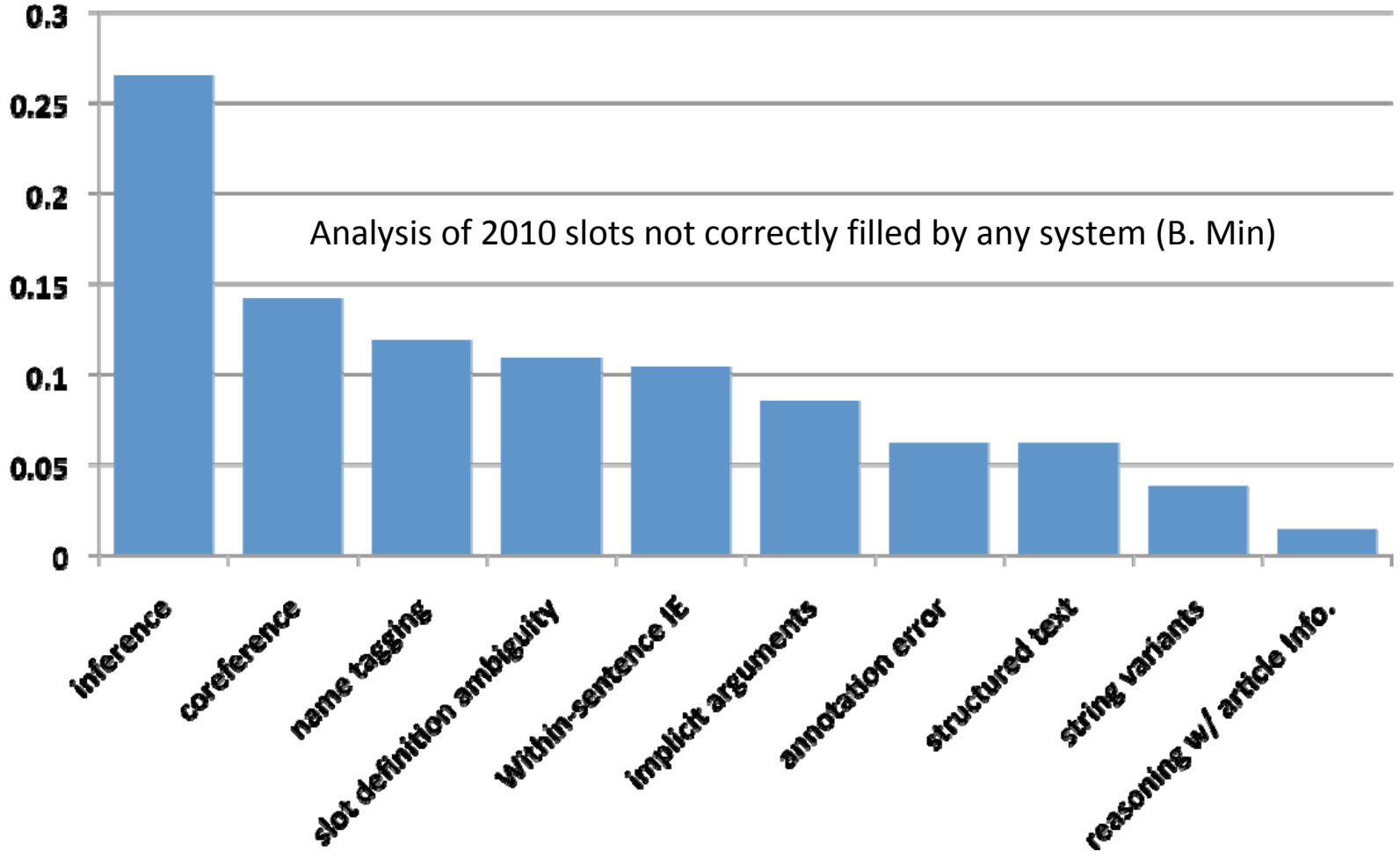

the ‘competition’ was stronger last year:

| | slots filled | distinct fills |
|---|---|---|
| 2010 | 623 | 1057 |
| 2011 | 498 | 953 |

# Performance without Document Validation



# Many Sources of Error



# IV: Temporal Slot Filling

# Temporal Slot Filling Task

- Many entity attributes such as a person's title and employer, and spouse change over time
- So we added a new task which requires that fills for selected slots be accompanied by time information. These *time intensive slots* are:
  - per:spouse
  - per:title
  - per:employee_of
  - per:member_of
  - per:cities_of_residence
  - per:stateorprovinces_of_residence
  - per:countries_of_residence
  - org:top_employees/members
- For the regular temporal task, slot fills and temporal information must be gathered across the entire corpus

- For the diagnostic temporal slot filling task, the system is given a correct slot fill and must extract the time information for that slot fill from a single document

# Temporal Representation

- Challenges:
  - want to be consistent with ‘data base’ approach of KBP
  - accommodate incomplete information
  - accommodate different granularities
- Solution:
  - express constraints on start and end times for slot value
  - 4-tuple $<t_1, t_2, t_3, t_4>$: $t_1 < t_{start} < t_2$ $t_3 < t_{end} < t_4$

| **Document text (2001-01-01)** | **T1** | **T2** | **T3** | **T4** |
|---|---|---|---|---|
| Chairman Smith | -infinite | 20010101 | 20010101 | +infinite |
| Smith, who has been chairman for two years | -infinite | 19990101 | 20010101 | +infinite |
| Smith, who was named chairman two years ago | 19990101 | 19990101 | 19990101 | +infinite |
| Smith, who resigned last October | -infinite | 20001001 | 20001001 | 20001031 |
| Smith served as chairman for 7 years before leaving in 1991 | 19840101 | 19841231 | 19910101 | 19911231 |
| Smith was named chairman in 1980 | 19800101 | 19801231 | 19800101 | +infinite |

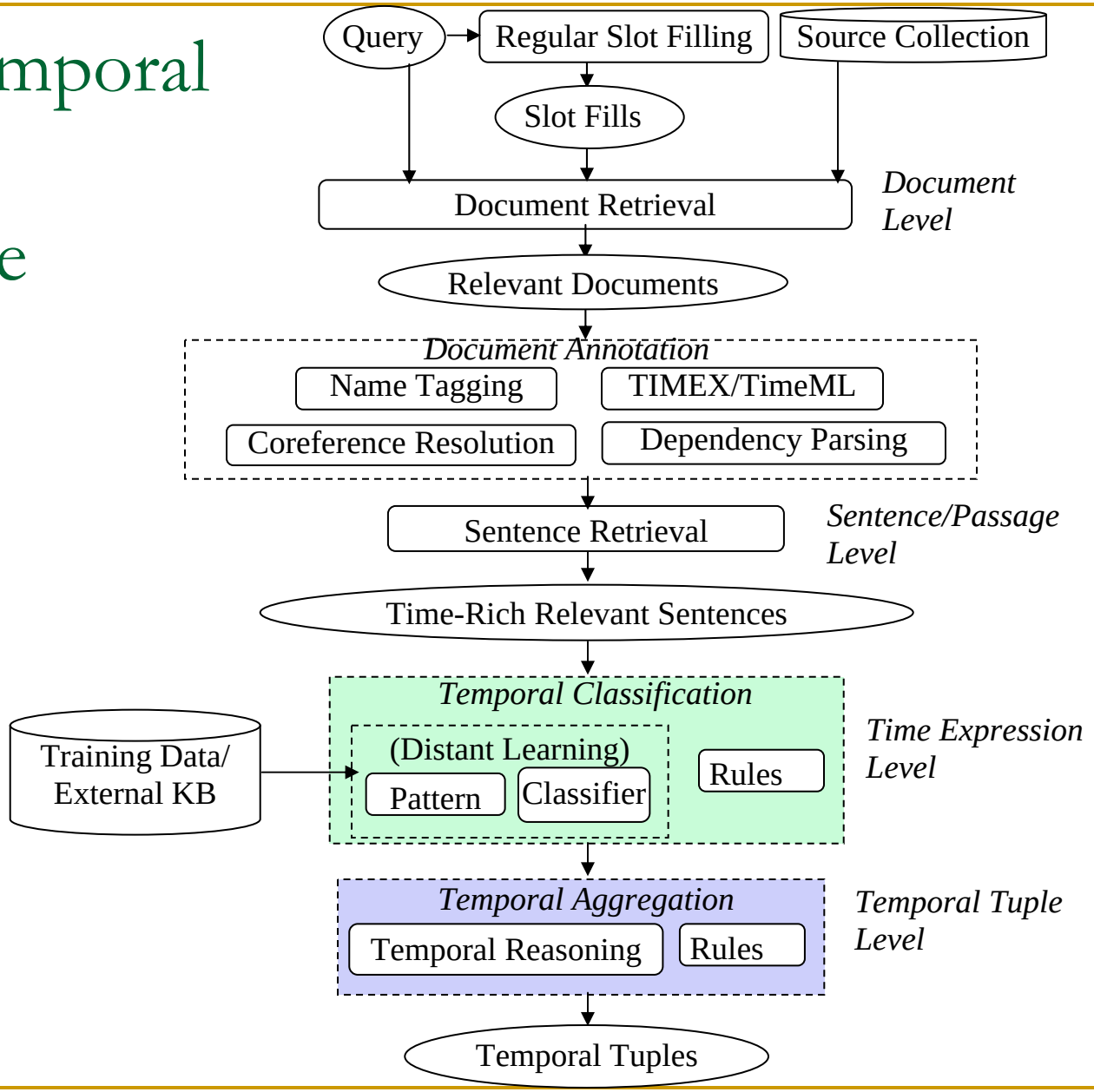
# Temporal Evaluation Metric

- New Evaluation Metric
  - Let <$t_1$, $t_2$, $t_3$, $t_4$> be system output,
    <$g_1$, $g_2$, $g_3$, $g_4$> be gold standard

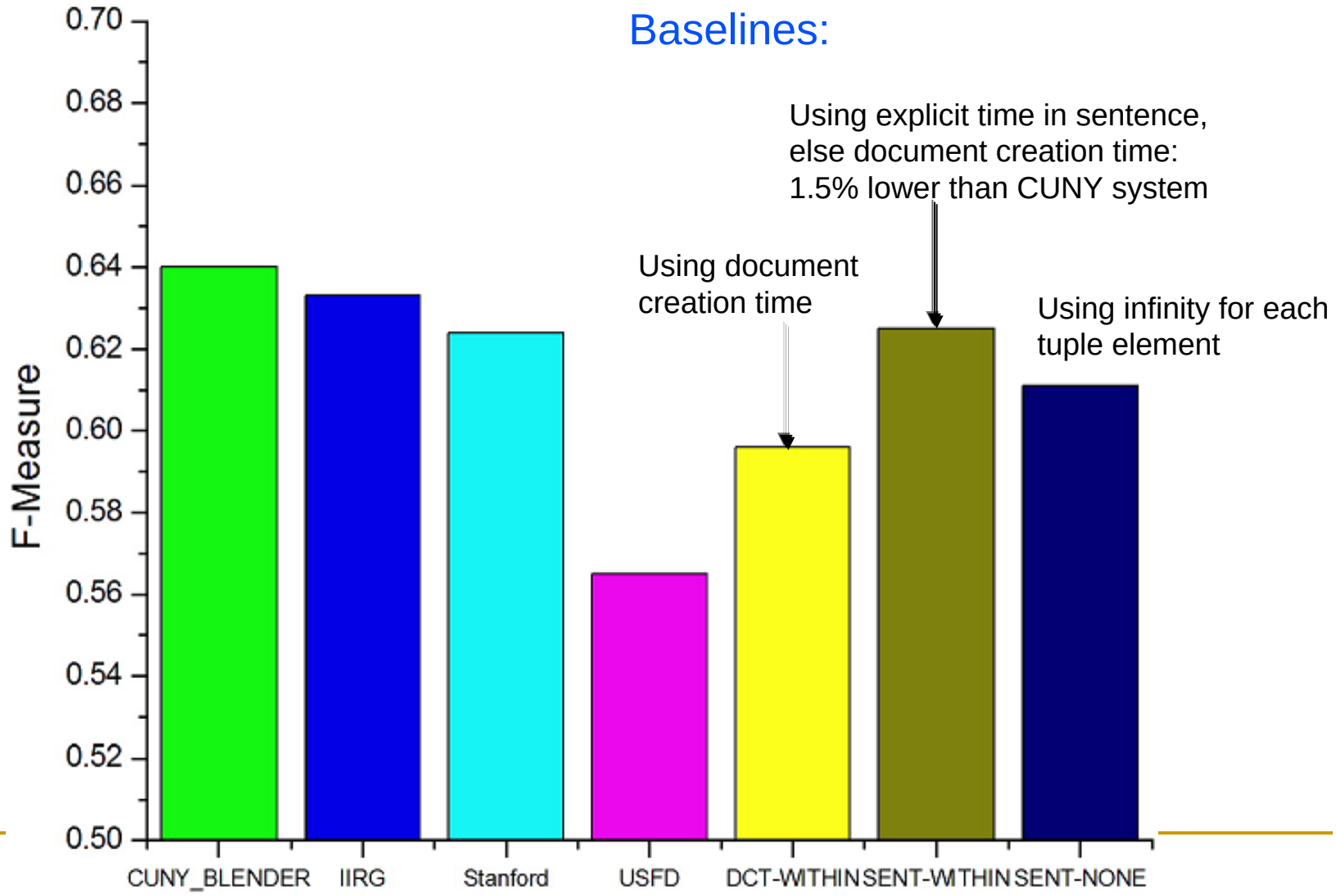
$$Q(S) = \frac{1}{4} \sum_i \frac{c}{c + |t_i - g_i|}$$

  - An error of *c* time units produces a 0.5 score
    - scores produced with c = 1 year
  - Each element in tuple is scored independently
  - For temporal SF task, a correct slot fill with temporal information ***t*** gets credit Q(S) (instead of 1)

# General Temporal SF System Architecture



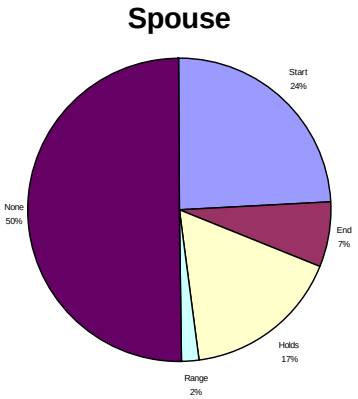
# Diagnostic System Performance

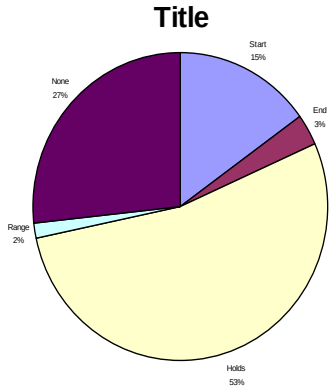


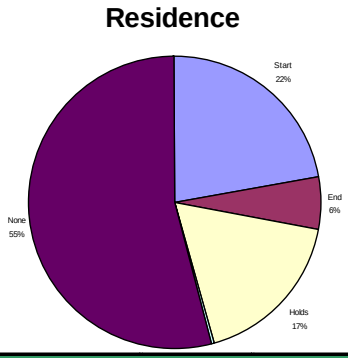
# But don't get too depressed yet…

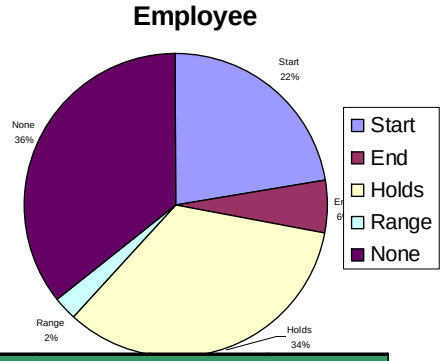
- Distant supervision data

| | Total | Start | End | Holds | Range | None |
|---|---|---|---|---|---|---|
| Spouse | 10196 | 2463 | 716 | 1705 | 182 | 5130 |
| Title | 14983 | 2229 | 501 | 7989 | 275 | 3989 |
| Employee | 17315 | 3888 | 965 | 5833 | 403 | 6226 |
| Residence | 4168 | 930 | 240 | 727 | 18 | 2253 |



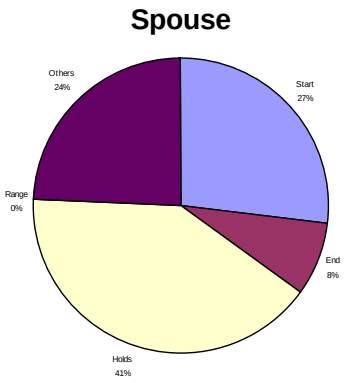



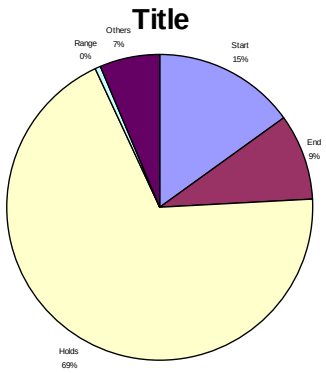



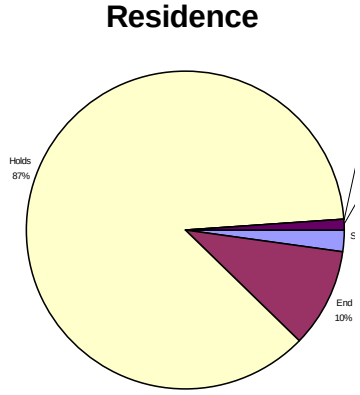



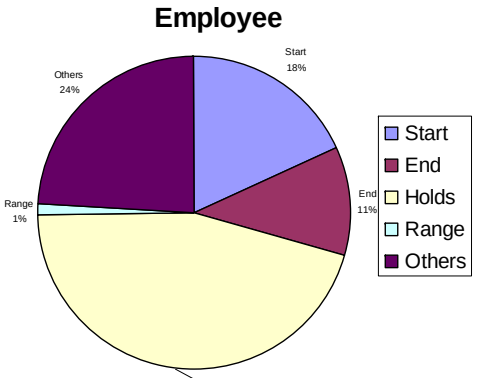

- KBP 2011 training data

| | Total | Start | End | Holds | Range | Others |
|---|---|---|---|---|---|---|
| **Spouse** | **28** | **10** | **3** | **15** | **0** | **9** |
| **Title** | **461** | **69** | **42** | **318** | **2** | **30** |
| **Employee** | **592** | **111** | **67** | **272** | **6** | **146** |
| **Residence** | **91** | **2** | **9** | **79** | **0** | **1** |



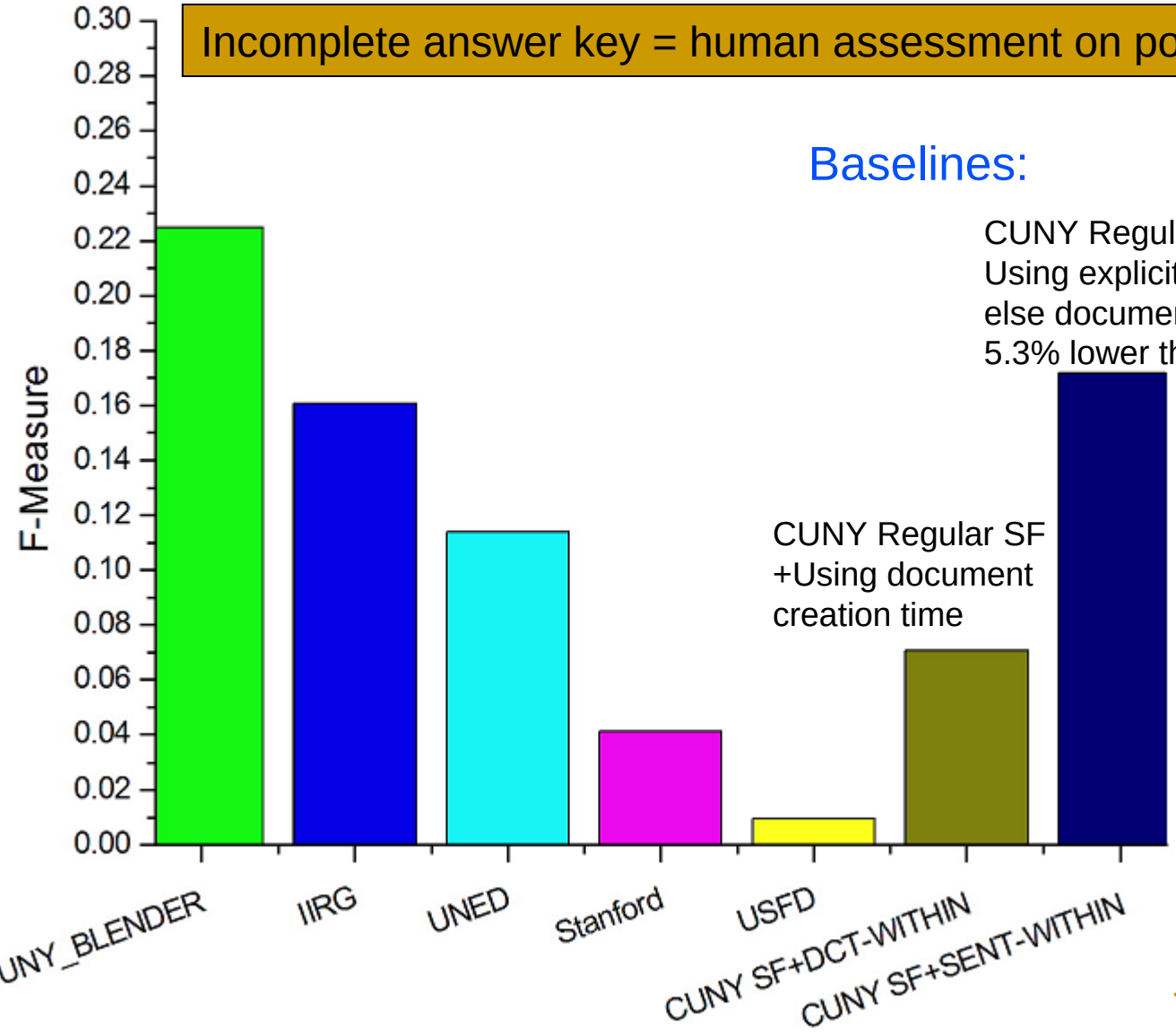



# Full System Performance: More Encouraging Results

Incomplete answer key = human assessment on pooled system output

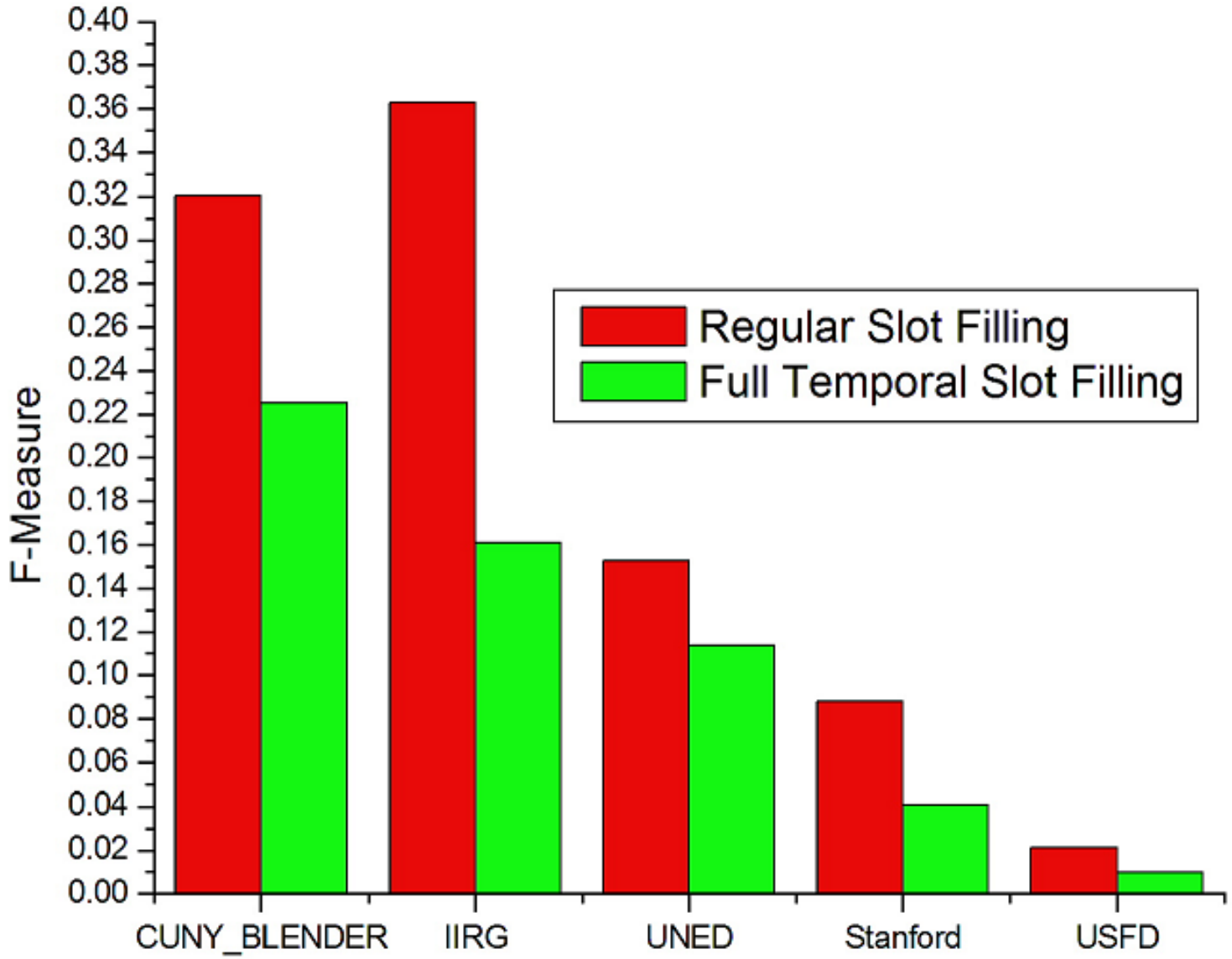



Baselines:

CUNY Regular SF +
Using explicit time in sentence, else document creation time:
5.3% lower than CUNY system

CUNY Regular SF +Using document creation time

# Impact of Regular SF on full TSF



# TSF Techniques

- What Works (Artiles et al., 2011; Li et al., 2011)
  - Enhance distant supervision through rich annotation, feature reduction and semi-supervised re-labeling
  - Combining flat approach and structured approach
  - Dynamically set time reference for text segment followed by a time expression

- Remaining Challenges
  - Implicit and wide context
  - Co-reference resolution errors
  - Temporal reasoning is needed for further improvement
  - Long-tail distribution of patterns

# Assessment and Prospects for 2012

- Mono-lingual Entity Linking
  - Approaches are converging
  - System performance on the basic task has continued to improve
    - the best systems are approaching human performance
  - NIL clustering successful
    - most cases in this year's evaluation could be handled by string matching alone
  - Is this task worth repeating?
    - more challenging cases for NIL clustering? extend to other genres?
  - Extend to Entity and Attribute Search?

- Cross-lingual Entity Linking
  - Overall performance only slightly lower than for the mono-lingual task
  - Person names and NIL clustering particularly challenging
  - New genres (web data, …)? New foreign languages (Arabic, …)?
  - Need another year for task to mature;  may want to
    - Provide more resources for Person name translation
    - Provide more training data for NIL clustering

# Assessment and Prospects For 2012

- Slot Filling
  - Seems hard to push above F = 0.30
    - low scores discourage publication
  - High entry cost for competitive performance
    - needs good NE, good coref, good syntactic analysis, …
    - makes it harder to evaluate more exotic approaches
    - failures scattered across modules ➔ must improve each module (expensive)
  - What might help?
    - fewer slots? richer annotation of training data? sharing more resources? focus on answer/passage validation? separate extraction and inference?
- Temporal Slot filling
  - very challenging – 2011 pilot helped to understand problems
    - need to select representative queries and documents
    - can we reduce burden of evaluation?
- Cross-lingual slot filling – a possibility for 2012
  - Ideal for participants who think regular slot filling is too easy
  - Pilot specifications and annotation done this year
  - Will need to:
    - Design diagnostic tasks
    - Provide intermediate resources including name translation, answer validation, etc.

Thank you and Join KBP2012!