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
Steve Jobs, Apple founder, dies

October 05, 2011 | By Brandon Griggs, CNN

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.
Overview of KBP Tasks

- Documents in T
  - Mono-lingual Entity Linking
  - Mono-lingual Slot Filling
  - Slot Fills in T
  - Temporal Slot Filling
  - Temporal Tuples

- Documents in S
  - Cross-lingual Entity Linking
  - Cross-lingual Slot Filling

Entity Clusters in T

Knowledge Base in T
### 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

<table>
<thead>
<tr>
<th>Task Participants/Year</th>
<th>Entity Linking</th>
<th>Slot Filling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mono-lingual</td>
<td>Cross-lingual</td>
</tr>
<tr>
<td></td>
<td>Regular</td>
<td>Optional</td>
</tr>
<tr>
<td>#Teams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>2010</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>2011</td>
<td><strong>22</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>#Submissions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>35</td>
<td>-</td>
</tr>
<tr>
<td>2010</td>
<td>46</td>
<td>20</td>
</tr>
<tr>
<td>2011</td>
<td><strong>53</strong></td>
<td>15</td>
</tr>
</tbody>
</table>
I: Mono-lingual Entity Linking
Entity Linking: Create Wiki Entry?

Query type: persons, GPEs, organizations

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

\[
G(e, e') = \begin{cases} 
1 & \text{iff } L(e) = L(e') \land C(e) = C(e') \land GI(e) = SI(e) = GI(e') = SI(e') \\
0 & \text{otherwise}
\end{cases}
\]

\[
\text{Precision } B-Cubed^+ = \text{Avg}_e [\text{Avg}_{e'.C(e)=C(e')}[G(e, e')]]
\]

\[
\text{Recall } B-Cubed^+ = \text{Avg}_e [\text{Avg}_{e'.L(e)=L(e')}[G(e, e')]]
\]
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, DMI_R_INESCID)

- Query Classification
  - DMI_R_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

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

- Simple methods work reasonably well

Name String Matching

B-cubed F-Measure

- NIL
- Non-NIL
- All

Projects:
- LCC
- MS
- NUS
- Shime
- CUNY
- UIUC
- SRI
- COGCOMP
- CMRCR
- Stanford
- UBC
- BLENDEF
- HLT
- COE
- THUNLP
- DMIR
- INESC
- MSRA
- WBSG
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

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，即台南一中。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。

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

Chinese Queries
- Chinese Name
- Chinese Document
  - Name Translation
  - Machine Translation
    - English Name
    - English Document
      - English Mono-lingual Entity Linking
      - English KB
        - Exploit Cross-lingual KB Links
          - English Queries
            - Cross-lingual NIL Clustering
              - Final Answers
From Mono-lingual to Cross-lingual

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Task</th>
<th>All</th>
<th>NIL</th>
<th>Non-NIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity</td>
<td>Monolingual</td>
<td>12.9%</td>
<td>5.7%</td>
<td>9.3%</td>
</tr>
<tr>
<td></td>
<td>Cross-lingual</td>
<td>20.9%</td>
<td>14.0%</td>
<td>28.6%</td>
</tr>
</tbody>
</table>

B-cubed+ F-Measure

- Mono-lingual Queries
- Cross-lingual Queries
“何伯” (He Uncle) refers to “an 81-years old man” or “He Yingjie”

62%

“丰华中文学校 (Fenghua Chinese School)”

7.6%

“莱赫·卡钦斯基 (Lech Aleksander Kaczynsk) vs. 雅罗斯瓦夫·卡钦斯基 (Jaroslaw Aleksander Kaczynski)”

4.5%

News reporter “Xiaoping Zhang”,
Ancient people “Bao Zheng”

1.8%

1.4%

2.1%

1.1%

1.7%

12%

Information networks
Document-level context
Discourse reasoning
Background knowledge
No-clue entities
NIL singletons
Popularity-dominant entities
Name spelling
Surface context
Entity type
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

B-cubed F-Measure

- Mono-lingual
- Cross-lingual

LCC
CUNY_UIUC_SRI
HLTOE
HITS
CUNY_BLENDE
LSV
ECNU
SIEL_IITH
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
Jim Parsons, a graduate of the University of Houston School of Music, Theatre 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

Jim Parsons

<table>
<thead>
<tr>
<th>Born</th>
<th>James Joseph Parsons</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 24, 1973 (age 37)</td>
<td>Houston, Texas, U.S.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Actor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years active</td>
<td>2000–present</td>
</tr>
</tbody>
</table>
## Attribute Distribution in Regular Slot Filling

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

<table>
<thead>
<tr>
<th></th>
<th>slots filled</th>
<th>distinct fills</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>623</td>
<td>1057</td>
</tr>
<tr>
<td>2011</td>
<td>498</td>
<td>953</td>
</tr>
</tbody>
</table>
Performance without Document Validation

- Black line: Without Document Validation
- Red line: With Document Validation

F-Measure

- NYU
- IIIRG
- LSV
- Stanford
- PolyU
- COMP
- UAlberta
- PRIS
- ECNU
- CEA_LVIC
- USFD
- UNED
- ICL_KBP
Many Sources of Error

Analysis of 2010 slots not correctly filled by any system (B. Min)
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_{\text{start}} < t_2 \quad t_3 < t_{\text{end}} < t_4\>

<table>
<thead>
<tr>
<th>Document text</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chairman Smith</td>
<td>-infinite</td>
<td>20010101</td>
<td>20010101</td>
<td>+infinite</td>
</tr>
<tr>
<td>Smith, who has been chairman</td>
<td>-infinite</td>
<td>19990101</td>
<td>20010101</td>
<td>+infinite</td>
</tr>
<tr>
<td>for two years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith, who was named chairman</td>
<td>19990101</td>
<td>19990101</td>
<td>19990101</td>
<td>+infinite</td>
</tr>
<tr>
<td>two years ago</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith, who resigned last</td>
<td>-infinite</td>
<td>20001001</td>
<td>20001001</td>
<td>20001031</td>
</tr>
<tr>
<td>October</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith served as chairman for</td>
<td>19840101</td>
<td>19841231</td>
<td>19910101</td>
<td>19911231</td>
</tr>
<tr>
<td>7 years before leaving in</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith was named chairman in</td>
<td>19800101</td>
<td>19801231</td>
<td>19800101</td>
<td>+infinite</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
  - 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)

\[
Q(S) = \frac{1}{4} \sum_i \frac{c}{c + |t_i - g_i|}
\]
General Temporal SF System Architecture

Query \rightarrow \text{Regular Slot Filling} \rightarrow \text{Source Collection}

\text{Slot Fills} \rightarrow \text{Document Retrieval} \rightarrow \text{Relevant Documents}

\text{Document Annotation}:
- Name Tagging
- TIMEX/TimeML
- Coreference Resolution
- Dependency Parsing

\text{Sentence Retrieval} \rightarrow \text{Time-Rich Relevant Sentences}

\text{Training Data/External KB} \rightarrow \text{Temporal Classification (Distant Learning)}
- Pattern
- Classifier
- Rules

\text{Temporal Aggregation}:
- Temporal Reasoning
- Rules

\text{Temporal Tuples}
Diagnostic System Performance

Baselines:
- Using infinity for each tuple element
- Using document creation time
- Using explicit time in sentence, else document creation time: 1.5% lower than CUNY system

The graph shows the performance of different diagnostic systems, with 'CUNY_BLENDE' achieving the highest F-Measure of approximately 0.64, followed by 'IIRG', 'Stanford', 'USFD', and 'DCT-WITHIN-SENT-WITHIN-SENT-NONE'.
But don’t get too depressed yet…

- **Distant supervision data**

<table>
<thead>
<tr>
<th>Category</th>
<th>Total</th>
<th>Start</th>
<th>End</th>
<th>Holds</th>
<th>Range</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse</td>
<td>10196</td>
<td>2463</td>
<td>716</td>
<td>1705</td>
<td>182</td>
<td>5130</td>
</tr>
<tr>
<td>Title</td>
<td>14983</td>
<td>2229</td>
<td>501</td>
<td>7989</td>
<td>275</td>
<td>3989</td>
</tr>
<tr>
<td>Employee</td>
<td>17315</td>
<td>3888</td>
<td>965</td>
<td>5833</td>
<td>403</td>
<td>6226</td>
</tr>
<tr>
<td>Residence</td>
<td>4168</td>
<td>930</td>
<td>240</td>
<td>727</td>
<td>18</td>
<td>2253</td>
</tr>
</tbody>
</table>

- **KBP 2011 training data**

<table>
<thead>
<tr>
<th>Category</th>
<th>Total</th>
<th>Start</th>
<th>End</th>
<th>Holds</th>
<th>Range</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse</td>
<td>28</td>
<td>10</td>
<td>3</td>
<td>15</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Title</td>
<td>461</td>
<td>69</td>
<td>42</td>
<td>318</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Employee</td>
<td>592</td>
<td>111</td>
<td>67</td>
<td>272</td>
<td>6</td>
<td>146</td>
</tr>
<tr>
<td>Residence</td>
<td>91</td>
<td>2</td>
<td>9</td>
<td>79</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
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
Impact of Regular SF on full TSF

![Bar chart showing F-Measure for various entities with Regular Slot Filling and Full Temporal Slot Filling.

CUNY_BLENDER, IIRG, UNED, Stanford, USFD]
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!