From Strings to Things: KELVIN in 2016 KBP and EDL

Tim Finin¹, Dawn Lawrie², James Mayfield, Paul McNamee and Craig Harman

Human Language Technology Center of Excellence
Johns Hopkins University

¹University of Maryland, Baltimore County
²Loyola University Maryland

November 14, 2016
Kelvin

• **KELVIN**: Knowledge Extraction, Linking, Validation and Inference

• Developed at the *Human Language Technology Center of Excellence* at JHU and used in TAC KBP (2010-16), EDL (2015-16) and other projects

• Takes English, Chinese & Spanish documents and produce a knowledge base in several formats

• We’ll review its monolingual processing, look at the multi-lingual use case and briefly present and assess the 2016 results
NIST TAC
NIST Text Analysis Conference

Yearly evaluation workshops on natural language processing (NLP) and related applications with large test collections and common evaluation procedures

<table>
<thead>
<tr>
<th>TAC Tracks</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Answering</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Textual Entailment</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarization</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Knowledge Base Population</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Entity Discovery &amp; Linking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Event Detection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Belief/Sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

http://www.nist.gov/tac
Knowledge Base Population

Tracks focused on building a knowledge base from entities and relations extracted from text

- **Cold Start KBP**: construct a KB from text
- **Entity discovery & linking**: cluster and link entity mentions
  - Slot filling
  - Slot filler validation
- **Sentiment**
- **Events**: discover and cluster events in text
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that...
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that...
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that...
Cold Start

Schema
<table>
<thead>
<tr>
<th>Relation</th>
<th>Inverse(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>per:children</td>
<td>per:parents, per:other_family</td>
</tr>
<tr>
<td>per:other_family</td>
<td>per:parents, per:children</td>
</tr>
<tr>
<td>per:parents</td>
<td>per:other_family, per:children</td>
</tr>
<tr>
<td>per:siblings</td>
<td>per:siblings, per:spouse</td>
</tr>
<tr>
<td>per:spouse</td>
<td>per:spouse, {org,gpe}:employees*</td>
</tr>
<tr>
<td>per:employee_of</td>
<td>org:membership*</td>
</tr>
<tr>
<td>per:member_of</td>
<td>org:students*</td>
</tr>
<tr>
<td>per:schools_attended</td>
<td>gpe:births_in_city*</td>
</tr>
<tr>
<td>per:city_of_birth</td>
<td>gpe:births_in_stateorprovince*</td>
</tr>
<tr>
<td>per:country_of_birth</td>
<td>gpe:births_in_country*</td>
</tr>
<tr>
<td>per:stateorprovince_of_birth</td>
<td>gpe:residents_of_city*</td>
</tr>
<tr>
<td>per:stateorprovince_of_birth</td>
<td>gpe:residents_of_stateorprovince*</td>
</tr>
<tr>
<td>per:states_or_provinces_of_residence</td>
<td>gpe:residents_of_country*</td>
</tr>
<tr>
<td>per:country_of_residence</td>
<td>gpe:residents_of_country*</td>
</tr>
<tr>
<td>per:city_of_death</td>
<td>gpe:deaths_in_city*</td>
</tr>
<tr>
<td>per:stateorprovince_of_death</td>
<td>gpe:deaths_in_stateorprovince*</td>
</tr>
<tr>
<td>per:country_of_death</td>
<td>gpe:deaths_in_country*</td>
</tr>
<tr>
<td>org:shareholders</td>
<td>{per,org,gpe}:holds_shares_in*</td>
</tr>
<tr>
<td>org:founded_by</td>
<td>{per,org,gpe}:organizations_founded*</td>
</tr>
<tr>
<td>org:top_members_employees</td>
<td>per:top_member_employee_of*</td>
</tr>
<tr>
<td>{org,gpe}:member_of</td>
<td>org:members</td>
</tr>
<tr>
<td>org:members</td>
<td>{org,gpe}:member_of</td>
</tr>
<tr>
<td>org:parents</td>
<td>{org,gpe}:subsidiaries</td>
</tr>
<tr>
<td>org:subsidiaries</td>
<td>org:parents</td>
</tr>
<tr>
<td>org:city_of_headquarters</td>
<td>gpe:headquarters_in_city*</td>
</tr>
<tr>
<td>org:stateorprovince_of_headquarters</td>
<td>gpe:headquarters_in_stateorprovince*</td>
</tr>
<tr>
<td>org:country_of_headquarters</td>
<td>gpe:headquarters_in_country*</td>
</tr>
</tbody>
</table>
String-Filled Relations

per:alternate_names  org:alternate_names
per:date_of_birth    org:political_religious_affiliation
per:age             org:number_of_employees_members
per:origin          org:date_founded
per:date_of_death   org:date_dissolved
per:cause_of_death  org:website
per:title
per:religion
per:charges
Cold Start

Schema

per:children
per:other_family
per:parents
per:siblings
per:spouse
per:employee_of
per:member_of
per:schools_attended
per:city_of_birth
per:stateorprovince_of_birth
per:country_of_birth
per:cities_of_residence
per:statesorprovinces_of_residence
per:countries_of_residence
per:city_of_death
per:stateorprovince_of_death
per:country_of_death
org:shareholders
org:founded_by
When Lisa’s mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.
The Task

You are given:

Lisa
Marge Simpson
Rancho Relaxo
Fuzzy Snuggleduck

When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.
When Lisa’s mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.
How do you know that your KB is any good?
How do you know that your KB is any good? 

Align it to a ground truth KB

How do you know that your KB is any good? 

Align it to a ground truth KB

Springfield

Springfield Elementary

Lisa Simpson

Bart Simpson

Marge Simpson

Homer Simpson

Bottomless Pete, Nature’s Cruelest Mistake

per:alternate_names

per:children

per:children

per:spouse

per:schools_attended

per:cities_of_residence
How do you know that your KB is any good?

Align it to a ground truth KB

But how are you going to produce ground truth? And wouldn’t the alignment be intractable anyway if the KB were of any reasonable size?
Where did the children of Marge Simpson go to school?
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated European adult movies available on their cable channels.

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that...
# Sample Evaluation Queries

<table>
<thead>
<tr>
<th>Query Entity</th>
<th>First Relation</th>
<th>Second Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adriana Petryna</td>
<td>per:title</td>
<td></td>
</tr>
<tr>
<td>Blackstone Group</td>
<td>org:founded_by</td>
<td></td>
</tr>
<tr>
<td>William Shore</td>
<td>per:organizations_founded</td>
<td>org:date_founded</td>
</tr>
<tr>
<td>Wistar Institute</td>
<td>org:employees</td>
<td>per:title</td>
</tr>
<tr>
<td>Andrew W. Mellon</td>
<td>per:children</td>
<td>per:organizations_founded</td>
</tr>
<tr>
<td>Lycee Alliance Israelite Universelle</td>
<td>org:employees</td>
<td>per:schools_attended</td>
</tr>
<tr>
<td>Tsitsi Jaji</td>
<td>per:schools_attended</td>
<td>org:students</td>
</tr>
</tbody>
</table>
Knowledge Base Population

• Read 90K documents: newswire articles & social media posts in English, Chinese and Spanish
• Find entity mentions, types and relations
• Cluster entities within and across documents and link to a reference KB when appropriate
• Remove errors (Obama born in Illinois), draw sound inferences (Malia and Sasha sisters)
• Generate triples with provenance data (Doc + string offsets) for mentions and relations

* 2016
KBP Coldstart and EDL

Build a system that can

• Read 90K documents: newswire articles & social media posts in English, Chinese and Spanish
• Find entity mentions, types and relations
• Cluster entities within and across documents and link to a reference KB when appropriate
• Remove errors (Obama born in Illinois), draw sound inferences (Malia and Sasha sisters)
• Generate triples with provenance data (Doc + string offsets) for mentions and relations

<DOC id="APW_ENG_20100325.0021" type="story">
  <HEADLINE>
    Divorce attorney says Dennis Hopper is dying
  </HEADLINE>
  <DATELINE>
    LOS ANGELES 2010-03-25 00:15:51 UTC
  </DATELINE>
  <TEXT>
    Dennis Hopper's divorce attorney says in a court filing that the actor is dying and can't undergo chemotherapy as he battles prostate cancer.
    <P>
      Attorney Joseph Mannis described the "Easy Rider" star's grave condition in a declaration filed Wednesday in Los Angeles Superior Court.
    </P>
    <P>
      Mannis and attorneys for Hopper's wife Victoria are fighting over when and whether to take the actor's deposition.
    </P>
  </TEXT>
</DOC>
KBP Coldstart and EDL

Build a system that can

• Read 90K documents: newswire articles & social media posts in English, Chinese and Spanish
• Find entities, types and relations
• Cluster entities within and across documents and link to a reference KB when appropriate
• Remove errors (Obama born in Illinois), draw sound inferences (Malia and Sasha sisters)
• Generate triples with provenance data (Doc + string offsets) for entities and relations

<DOC id="APW ENG 20100325.0021" type="story">
<HEADLINE>Divorce attorney says Dennis Hopper is dying</HEADLINE>
<DATELINE>LOS ANGELES 2010-03-25 00:15:51 UTC</DATELINE>

Dennis Hopper's divorce attorney described in a court filing that the actor is dying and can't undergo chemotherapy.

A attorney Joseph Mannis described the "Easy Rider" star's grave condition in a declaration filed Wednesday in Los Angeles Superior Court.

Mannis and attorneys for Hopper's wife Victoria are fighting over when and whether to take the actor's deposition.

...
Divorce attorney says Dennis Hopper is dying

A\'s\’ attorney Joseph Mannis described the “Easy Rider” star\'s grave condition in a declaration filed Wednesday in Los Angeles Superior Court.

Mannis and attorneys for Hopper\'s wife Victoria are fighting over when and whether to take the actor\'s deposition.

Dennis Hopper\’s divorce attorney said in a court filing that the actor is dying and can\'t undergo chemotherapy as the battle prostate cancer.

Mannis and attorneys for Hopper\’s wife Victoria are fighting over when and whether to take the actor\'s deposition.

Dennis Hopper\’s divorce attorney said in a court filing that the actor is dying and can\'t undergo chemotherapy as the battle prostate cancer.

Mannis and attorneys for Hopper\’s wife Victoria are fighting over when and whether to take the actor\'s deposition.
KB Evaluation Methodology

• Evaluating KBs extracted from 90K documents is non-trivial

• TAC’s approach is simplified by:
  – Fixing the ontology of entity types and relations
  – Specifying a serialization as triples + provenance
  – Sampling a KB using a set of queries grounded in an entity mention found in a document

• Given a KB, we can evaluate its precision and recall for a set of queries
KB Evaluation Methodology

• **A query:** What are the names of schools attended by the children of the entity mentioned in document #45611 at characters 401-412
  – That mention is *George Bush* and the document context suggests it refers to the 41st U.S. president
  – Query given in structured form using TAC ontology

• **Assessors** determine good answers in corpus and check submitted results using their provenance

• **Answers:** entities for Yale, Harvard, Tulane, UT Austin, Univ. of Virginia, Boston College, ...
TAC Ontology

• Five basic entity types
  – PER: **people** (John Lennon) or groups (Americans)
  – ORG: **organizations** like IBM, MIT or US Senate
  – GPE: **geopolitical** entity like Boston, Belgium or Europe
  – LOC: **locations** like Lake Michigan or the Rockies
  – FAC: **facilities** like BWI or the Empire State Building

• Entity Mentions
  – **Strings** referencing entities by name (Barack Obama), description (the President) or pronoun (his)

• ~65 relations
  – Relations hold between two entities: parent_of, spouse, employer, founded_by, city_of_birth, ...
  – Or between an entity & string: age, website, title, cause_of_death, ...
Our ontology has official TAC types/relations and many more we capture from tools and infer from the data
Monolingual Kelvin

1. Document level NLP and information extraction
2. Integrate, fix problems, map to TAC ontology and requirements, draw inferences, etc.
3. Cross-document co-reference to form an initial KB
4. KB-level processing, inference, entity linking, additional co-ref, adjudication, ...
5. Materialize KBs: TAC, EDL, RDF, etc.

IE
TAC
CR
KB
MAT

documents

KBs
1 Initial NLP/IE Processing

- Process documents in parallel on a grid, applying a number of tools to find mentions, entities, relations and events
- In 2016 we used BBN’s Serif and Facets but have also used CoreNLP and plan to use others
- We produce an Apache Thrift object for each document that includes the text and the relevant data produced by the tools using the common Concrete schema
Concrete

• HLTCOE Apache Thrift schema for NLP data
• Available at http://hltcoe.github.io/
• Common objects that many tools can use and contribute to
  – Serif, Facets, CoreNLP, open IE, relation extraction, sentiment analytics, translation, language ID, etc.
  – Visualization, annotation, etc.
• API bindings for Java, Python, Clojure, Perl, Ruby, etc.

Concrete schema types

• audio
• cluster
• communication
• email
• entities
• ex
• language
• linking
• metadata
• nitf
• services
• situations
• spans
• structure
• twitter
• uuid
Quicklime Visualization

Web-based Concrete visualization and annotation tool

• Constituency parse trees
• Intra-doc co-reference
• Dependency parse graphs
• NER tags
• POS tags
2 Integrating the NLP data

Process Concrete objects in parallel to:

- Integrate data from tools (e.g., Stanford and Serif)
- Fix problems, e.g., trim mention strings, find missed mentions, deconflict tangled mention chains, etc.
- Select canonical mentions
- Map types and relations to official TAC ontology and an extended COE ontology
- Extract relations from events (life.born => date and place of birth)
- Add mentions from tags such as <post_author>
3 Kripke: Cross-Doc Coref

- Cross-document co-reference creates an initial KB from the collection of single-document KBs
  - Identify that Barack Obama entity in DOC32 is same individual as Obama in DOC342, etc.
- Language agnostic; works well for ENG, CMN, SPA document collections
- Only uses entity name mention strings
  - Thus supports a wide range of applications
  - Option to also use nominal mentions
3 Kripke: Cross-Doc Coref

- Untrained, agglomerative clustering
  - Document entities start as singleton clusters
  - Combine if type match, good name match & good context match
  - Cascade of merging steps w. relaxing constraints, no splits
  - Maintain index of name variants \(\rightarrow\) cluster

- Context: bag of entity name mentions
  - No use of use document text or relations
  - Does not link entities to a background KB
3 Kripke: Cross-Doc Coref

- 30K ENG docs: Kripke mapped 430K document entities into
  - 210K KB entities using name mentions
  - NNN KB entities if also using nominals
- Conservative, tending to under- rather than over-merge
- Typical pattern for a popular entity is one large cluster and a long tail of small ones
  - Barack Obama had one cluster of 8K doc. entities and ~100 small ones
4 KB level processing

- After step 3 we have a single, unified knowledge base
- We go through a number of steps involving entity linking, entity merging, inference, and slot value consolidation
Inference and adjudication rules used to

- Delete relations violating ontology constraints
  - *Person can’t be born in an organization*
  - *Person can’t be her own parent or spouse*

- Add relations from sound inference rules
  - *Two people sharing a parent are siblings*
  - *X born in place P₁, P₁ part of P₂ => X born in P₂*

- Add relations by default or heuristic rules
  - *Person probably citizen of their country of birth*
Entity Linking

- Simple strategy links entities to reference KB
- TAC uses last Freebase KB, our subset has
  - ~4.5M entities and ~150M triples
  - Names and text in English, Spanish and Chinese
- Compare entity mentions with Freebase entity names & aliases, weighted by
  - How often each mention used for entity
  - Entity significance (log of inlinks (1..20))
- Results: set of candidates and score for each
  - Reject candidates if too many or scores low
KB-level merging rules

- Merge cities with same name in same state
- Merge entities of same type linked to same Freebase entity
- Highly discriminative relations give evidence of sameness
  - per:spouse is few to few
  - org:top_level_employee is few to few
- Merge PERs with similar names who were
  - Both married to the same person, or
  - Both CEOs of the same company, or ...
Slot Value Consolidation

• **Problem:** too many values for some slots, especially for ‘popular’ entities, e.g.
  – An entity with four different *per:age* values
  – Obama has >100 *per:employee_of* values

• **Strategy:** rank values and select best
  – Rank attested values by # of attesting docs
  – Rank inferred values below attested values
  – Select best value for single-values slot
  – Select best N values for multi-valued slots, where N depends on slot
Slot Value Consolidation

# of attesting documents for slots with many values follows a power law, as in Obama’s employers from CS’13

Multi-valued slots use two thresholds for accepting nth candidate
- accept if \( n \leq T_1 \)
- accept \( n \leq T_2 \) and \( >1 \) attesting document

<table>
<thead>
<tr>
<th>predicate</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>per:children</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>per:countries_of_residence</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>per:employee_of</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>per:religion</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>per:schools_attended</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>per:spouse</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>
Inference in Cold Start

- Can only use “sound” inference rules, e.g.
  - People are siblings if they share a parent
  - People born in city X were born in region Y if X is in Y
- Default & heuristic rules not attested
  - You probably resided in city of your school
- 2014 base run found 848 attested sibling relations, inference added 132, 16% more
- Drawing sound inferences from \{good|bad\} facts yields \{good|bad\} results
Materialize KB versions

Encode our nascent KB in one or more actual KB systems or serialization

- TAC has custom serialization, e.g., at most 4 provenance strings of length ≤ 150 chars
  - Pick best for slots with many provenance strings
- EDL has custom serialization
- RDF requires OWL schema & reification for metadata, e.g., provenance
Multi-lingual Kelvin
Multilingual KBP

• Many examples where facts from different languages combine to answer queries or support inference

  Q: Who lives in the same city as Bodo Elleke?
  A: Frank Ribery aka Franck Ribéry aka 里贝里

• Why we know both live in Munich:

     ...said the younger Bodo Elleke, who was born in Schodack in 1930 and is now a retired architect who lives in Munich.

     拉霍伊在接受西班牙国家电台的采访时肯定，今年的三位金球奖热门候选人中，梅西“度过了一个出色的赛季”，而拜仁慕尼黑球员里贝里则“赢得了一切”

• Kripke merged entities with mentions Frank Ribery, Franck Ribéry & 里贝里
Monolingual to Multilingual Kelvin

Zoom in on our cross-doc co-ref step

- Concatenate document-level KBs to form a **DOC KB** as input to Kripke
- Kripke outputs a set of **CLUSTERS** defining an equivalence relation
- Merger uses **CLUSTERS** to combine **DOC KB** entities, yielding the initial KB
- We use the **DOC KB** and **CLUSTERS** from each language to create an initial multilingual KB
Trilingual KBP & EDL

• Kripke computes CLUSTERS for combined multilingual DOC KBS
Translating non-English mentions to English, when possible enhances clustering.

- Kripke computes CLUSTERS for combined monolingual DOC KBS
- Optionally translate non-English mentions to English, when possible enhances clustering

Trilingual KBP & EDL
• Kripke computes CLUSTERS for combined monolingual DOC KBS
• Optionally translate non-English mentions
• Use all four CLUSTERS to merge entities in the three DOC KBS
Results
2016 TAC KBP Results

• For the 7 KB and 11 SF submissions, depending on metric (macro/micro avg), we placed
  – 1\textsuperscript{st} or 2\textsuperscript{nd} of 5 on XLING and were the only team to do all three languages
  – 2\textsuperscript{nd} or 4\textsuperscript{th} of 18 on ENG depending on metric
  – 1\textsuperscript{st} or 2\textsuperscript{nd} of 4 on CMN depending on metric
  – We did poorly on SPA, finding few relations

• See workshop paper for details

• TAC EDL results are forthcoming
# 2016 TAC KBP Results

<table>
<thead>
<tr>
<th>Run</th>
<th>0-hop</th>
<th>1-hop</th>
<th>All-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GT</td>
<td>R</td>
<td>W</td>
</tr>
<tr>
<td>E1</td>
<td>801</td>
<td>194</td>
<td>367</td>
</tr>
<tr>
<td>E2</td>
<td>801</td>
<td>185</td>
<td>313</td>
</tr>
<tr>
<td>E3</td>
<td>801</td>
<td>186</td>
<td>311</td>
</tr>
<tr>
<td>E4</td>
<td>801</td>
<td>186</td>
<td>393</td>
</tr>
<tr>
<td>E5</td>
<td>801</td>
<td>195</td>
<td>366</td>
</tr>
<tr>
<td>C1</td>
<td>751</td>
<td>141</td>
<td>125</td>
</tr>
<tr>
<td>C2</td>
<td>751</td>
<td>141</td>
<td>125</td>
</tr>
<tr>
<td>C3</td>
<td>751</td>
<td>141</td>
<td>125</td>
</tr>
<tr>
<td>C4</td>
<td>751</td>
<td>143</td>
<td>127</td>
</tr>
<tr>
<td>S1</td>
<td>332</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>S2</td>
<td>332</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>S3</td>
<td>332</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>S4</td>
<td>332</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>X1</td>
<td>4094</td>
<td>838</td>
<td>1177</td>
</tr>
<tr>
<td>X2</td>
<td>4094</td>
<td>713</td>
<td>1107</td>
</tr>
<tr>
<td>X3</td>
<td>4094</td>
<td>697</td>
<td>936</td>
</tr>
<tr>
<td>X4</td>
<td>4094</td>
<td>321</td>
<td>385</td>
</tr>
</tbody>
</table>

Ground-truth, right, wrong & duplicate answers for 2016 KBP KB runs
## 2016 TAC KBP Results

| Run | 0-hop | | | 1-hop | | | All-hop | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
|     | P     | R     | F1    | P     | R     | F1    | P     | R     | F1    |
| E1  | 0.3458 | 0.2422 | 0.2849 | 0.0877 | 0.1152 | 0.0996 | 0.2197 | 0.1993 | 0.2090 |
| E2  | 0.3715 | 0.2310 | 0.2848 | 0.1033 | 0.1005 | 0.1019 | 0.2525 | 0.1869 | 0.2148 |
| E3  | 0.3742 | 0.2322 | 0.2866 | 0.1017 | 0.1005 | 0.1011 | 0.2522 | 0.1878 | 0.2153 |
| E4  | 0.3212 | 0.2322 | 0.2696 | 0.0417 | 0.0980 | 0.0585 | 0.1469 | 0.1869 | 0.1645 |
| E5  | 0.3476 | 0.2434 | 0.2863 | 0.0877 | 0.1152 | 0.0996 | 0.2206 | 0.2002 | 0.2099 |
| C1  | 0.5301 | 0.1877 | 0.2773 | 0.2984 | 0.2478 | 0.2708 | 0.4333 | 0.2018 | 0.2754 |
| C2  | 0.5301 | 0.1877 | 0.2773 | 0.2984 | 0.2478 | 0.2708 | 0.4333 | 0.2018 | 0.2754 |
| C3  | 0.5301 | 0.1877 | 0.2773 | 0.2984 | 0.2478 | 0.2708 | 0.4333 | 0.2018 | 0.2754 |
| C4  | 0.5296 | 0.1904 | 0.2801 | 0.2908 | 0.2478 | 0.2676 | 0.4292 | 0.2039 | 0.2764 |
| S1  | 0.3077 | 0.0120 | 0.0232 | 0.0000 | 0.0000 | 0.0000 | 0.3077 | 0.0073 | 0.0143 |
| S2  | 0.3077 | 0.0120 | 0.0232 | 0.0000 | 0.0000 | 0.0000 | 0.3077 | 0.0073 | 0.0143 |
| S3  | 0.5000 | 0.0090 | 0.0178 | 0.0000 | 0.0000 | 0.0000 | 0.5000 | 0.0055 | 0.0109 |
| S4  | 0.3077 | 0.0120 | 0.0232 | 0.0000 | 0.0000 | 0.0000 | 0.3077 | 0.0073 | 0.0143 |
| X1  | 0.4159 | 0.2047 | 0.2743 | 0.1521 | 0.1786 | 0.1643 | 0.2751 | 0.1962 | 0.2291 |
| X2  | 0.3918 | 0.1742 | 0.2411 | 0.1853 | 0.1384 | 0.1585 | 0.2996 | 0.1626 | 0.2108 |
| X3  | 0.4268 | 0.1702 | 0.2434 | 0.2054 | 0.1344 | 0.1625 | 0.3293 | 0.1586 | 0.2141 |
| X4  | 0.4547 | 0.0784 | 0.1337 | 0.1779 | 0.0539 | 0.0828 | 0.3280 | 0.0705 | 0.1160 |

Micro precision, recall and F1 scores for 2016 KBP KB runs.
## 2016 EDL XLING Results

<table>
<thead>
<tr>
<th></th>
<th>NER</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre</td>
<td>rec</td>
<td>F1</td>
<td>pre</td>
<td>rec</td>
<td>F1</td>
<td>pre</td>
<td>rec</td>
<td>F1</td>
<td>pre</td>
<td>rec</td>
<td>F1</td>
</tr>
<tr>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.656</td>
<td>0.573</td>
<td>0.612</td>
<td>0.489</td>
<td>0.427</td>
<td>0.456</td>
<td>0.368</td>
<td>0.614</td>
<td>0.460</td>
<td>0.470</td>
<td>0.411</td>
<td>0.438</td>
</tr>
<tr>
<td>2</td>
<td>0.656</td>
<td>0.573</td>
<td>0.612</td>
<td>0.488</td>
<td>0.426</td>
<td>0.455</td>
<td>0.335</td>
<td>0.624</td>
<td>0.436</td>
<td>0.469</td>
<td>0.410</td>
<td>0.437</td>
</tr>
<tr>
<td>3</td>
<td>0.656</td>
<td>0.573</td>
<td>0.612</td>
<td>0.476</td>
<td>0.416</td>
<td>0.444</td>
<td>0.346</td>
<td>0.615</td>
<td>0.443</td>
<td>0.457</td>
<td>0.399</td>
<td>0.426</td>
</tr>
<tr>
<td>4</td>
<td>0.656</td>
<td>0.573</td>
<td>0.612</td>
<td>0.489</td>
<td>0.427</td>
<td>0.456</td>
<td>0.335</td>
<td>0.625</td>
<td>0.436</td>
<td>0.469</td>
<td>0.410</td>
<td>0.438</td>
</tr>
<tr>
<td>5</td>
<td>0.661</td>
<td>0.563</td>
<td>0.608</td>
<td>0.494</td>
<td>0.420</td>
<td>0.454</td>
<td>0.374</td>
<td>0.612</td>
<td>0.464</td>
<td>0.474</td>
<td>0.404</td>
<td>0.436</td>
</tr>
</tbody>
</table>

XLING run precision, recall and F1 measures for four key metrics: strong typed mention match (NER), strong all match (Linking), strong nil match (Nil), and mention ceaf plus (Clustering)
2016 Results Observations

• Overall XLING1 was best
• Variations for monolingual runs were similar
  – Using translated mentions for non-English helped
  – Using nominal mentions seemed to improve cross-doc co-ref slightly
• EDL scores (and maybe KBP) lowered by bug in our nominal mention trimming code; the nominal strings correctly identified but offsets were wrong 😞
Kelvin Docker Container

• **Problem:** Kelvin is a large and complex system that’s difficult to port to a new Unix environment, let alone a different OS

• **Solution:** We use Docker to virtualize Kelvin as several containers that can be run on any system that supports Docker
  – e.g., most Unix systems, Mac OSX and Windows
Conclusion
Lessons Learned

• We always have to mind precision & recall
• Extracting information from text is inherently noisy; reading more text helps both
• Using machine learning at every level is important
• Making more use of probabilities will help
• Extracting information about a events is hard
• Recognizing the temporal extent of relations is important, but still a challenge
Conclusion

• KBs help in extracting information from text
• The information extracted can update the KBs
• The KBs provide support for new tasks, such as question answering and speech interfaces
• We’ll see this approach grow and evolve in the future
• New machine learning frameworks will result in better accuracy