Entity Linking and Slot Filling through Statistical Processing and Inference Rules

Dan Bikel, Vittorio Castelli, Radu Florian, Ding-Jung Han

November 16th, 2009
Overview

Information Extraction

Entity Linking

Slot Filling
ABBOTT PARK, Illinois

Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (euro645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago. Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Information Extraction

ABBOTT PARK, Illinois

Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago. Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Information Extraction

ABBOTT PARK, Illinois

Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Abbott Park, Illinois

Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Abbott Park, Illinois

Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million ( euro645 million ), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million, or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Information Extraction: Highlights

- **Language-independent infrastructure, based on a Markov model maximum entropy model**

- **Identifies:**
  - 40 mention types
  - 50 relation types
  - Named, nominal and pronominal coreference
Progress

Entity Linking

Slot Filling

Information Extraction
Entity Linking: Overall Approach

**Approach**

- Name (PER, ORG, GPE)
- Document with mention

- Fast match
- Detailed Match

- Success
- Failure
- NULL
- Best Match

**Much of the work: how to create and use database for good fast match**
Creating db: sources

- Gather information from two sources, knowledge base and DBpedia

### TAC Knowledge Base

- extracted by LDC from a Wikipedia dump

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>PER</td>
<td>114,523</td>
</tr>
<tr>
<td>ORG</td>
<td>55,813</td>
</tr>
<tr>
<td>GPE</td>
<td>116,498</td>
</tr>
<tr>
<td>UKN</td>
<td>531,907</td>
</tr>
<tr>
<td>Total</td>
<td>818,741</td>
</tr>
</tbody>
</table>

### DBpedia

- v3.2, which contains an ontology

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>PER</td>
<td>211,029</td>
</tr>
<tr>
<td>ORG</td>
<td>75,627</td>
</tr>
<tr>
<td>GPE</td>
<td>179,842</td>
</tr>
<tr>
<td>Total</td>
<td>466,498</td>
</tr>
</tbody>
</table>
Creating db: merging

- All but ~60k entities from DBpedia are already in KB

- Coerce types for UKN entities using DBpedia
  - 153k/532k UKN entities type-coerced

- Merge infobox slots obtained from DBpedia’s infoboxes dataset into KB entities’ slots
Creating db: search engine

- Lucene
- Each entity is a “document” in Lucene index
  - index
  - character trigrams
  - words in context
- Search function: based on character trigrams
- Add weights to entities corresponding to their “fame”
Detailed match for names

- **Name similarity score**
  - Investigated a large variety of matching functions
  - Use “Soft TF-IDF” string matching from SecondString library (Cohen *et al.*, 2003)
  - String similarity trained on sample of named entities
  - Scoring is “soft”
Health-care products maker Abbott Laboratories Inc. said Tuesday that its first-quarter earnings rose 1.8 percent from a year ago, as sales jumped 16 percent on strong contributions from branded products.

Net income increased to $837.9 million (€645 million), or 53 cents per share, in the three months ended March 31 from $822.9 million, or 52 cents per share, a year ago.

Excluding one-time items, earnings from continuing operations totaled 58 cents per share.
Detailed match for context

Query Context

Abbott Park  Illinois

Health-care products  maker  Abbott Laboratories Inc.  its
Detailed match for context

Query Context

Abbot Park  Illinois  Health-care products  maker
Abbott Laboratories Inc.  its

DB Context

Abbott Laboratories
A promise for life
1888
Abbott Park,   North Chicago, Illinois , USA
Miles D. White , Chairman and CEO

... Pharmaceutical products, medical devices, diagnostic assays</fact>
www.abbott.com/
Detailed match for context

**Query Context**
- Abbot Park
- Illinois
- Health-care products
- maker
- Abbott Laboratories Inc.
- its

**DB Context**
- 1888
- Abbott Laboratories
- A promise for life
- Abbott Park
- North Chicago
- Illinois
- USA
- medical devices
- diagnostic assays
- www.abbott.com
- Pharmaceutical products

**Context similarity score:**
- word inclusion of the query context in the DB context
Detailed match: overall similarity score

- **Overall similarity score**

\[
\text{Sim} (e_1, e_2) = \text{NameSim} (e_1, e_2) \cdot \text{ContextSim} (e_1, e_2)^\alpha
\]
Entity Linking Results

![Entity Linking Results Chart]

- Devtest:
  - KB: 92.1
  - NIL: 93.9
  - Total: 92.6

- Eval:
  - KB: 63.5
  - NIL: 71.8
  - Total: 68.2

**Pie Charts**

- Devtest Pie Chart
  - KB: 92.1%
  - NIL: 93.9%
  - Total: 92.6%

- Eval Pie Chart
  - KB: 63.5%
  - NIL: 71.8%
  - Total: 68.2%
Progress

Slot Filling

Information Extraction

Entity Linking
Slot Filling Task

**Input:**
- A name (GPE, PER, ORG)
- A DB id (optional)
- Document containing the name

**Output:**
- Fillers for all applicable slots

<query id="SF1">
  <name>Barack Obama</name>
  <docid>SUN-009</docid>
  <enttype>PER</enttype>
  <nodeid>E0566375</nodeid>
  <ignore>per:date_of_death per:place_of_death</ignore>
</query>
## Slots

<table>
<thead>
<tr>
<th>Person</th>
<th>Organization</th>
<th>Geo-Political Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>per:alternate_names</td>
<td>org:alternate_names</td>
<td>gpe:alternate_names</td>
</tr>
<tr>
<td>per:birthdate</td>
<td>org:political/religious_affiliation</td>
<td>gpe:capital</td>
</tr>
<tr>
<td>per:age</td>
<td>org:top_members/employees</td>
<td>gpe:subsidiary_orgs</td>
</tr>
<tr>
<td>per:place_of_birth</td>
<td>org:top_members/employees</td>
<td>gpe:top_employees</td>
</tr>
<tr>
<td>per:origin</td>
<td>org:members</td>
<td>gpe:political_parties</td>
</tr>
<tr>
<td>per:place_of_death</td>
<td>org:member_of</td>
<td>gpe:established</td>
</tr>
<tr>
<td>per:cause_of_death</td>
<td>org:subordinates</td>
<td>gpe:population</td>
</tr>
<tr>
<td>per:residences</td>
<td>org:parents</td>
<td>gpe:currency</td>
</tr>
<tr>
<td>per:schools_attended</td>
<td>org:found_by</td>
<td></td>
</tr>
<tr>
<td>per:title</td>
<td>org:found</td>
<td></td>
</tr>
<tr>
<td>per:member_of</td>
<td>org:found</td>
<td></td>
</tr>
<tr>
<td>per:employee_of</td>
<td>org:disolved</td>
<td></td>
</tr>
<tr>
<td>per:religion</td>
<td>org:headquarters</td>
<td></td>
</tr>
<tr>
<td>per:spouse</td>
<td>org:shareholders</td>
<td></td>
</tr>
<tr>
<td>per:children</td>
<td>org:website</td>
<td></td>
</tr>
<tr>
<td>per:parents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per:siblings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per:other_family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per:charges</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Slot Filling: Approach

- **Problem:** no annotated data to train a model
- **Solution:** build a hybrid system
  - Start with automatically extracted relations
  - Construct slots using these relations and hand-built inference rules (Horn clauses)
Relation Inference

per:date_of_birth(X,Y) :- bornOn(X,Y).
per:age(X,Y) :- ageOf(X,Y).
per:employee_of(X,Y) :- employedBy(X,Y).

per:religion(X,Y) :- partOfMany(X,Y), religious(Y).
per:religion(X,Y) :- located(X,Z), religiousFacility(Z,Y).

per:origin(X,Y) :- isOrigin(Y), coref(X,Y).
per:origin(X,Y) :- partOfMany(X,Y), isGPE(Y).
per:origin(X,Y) :- per:sibling(X,Z), per:origin(Z,Y).
per:origin(X,Y) :- partOfMany(X,Z), per:origin(Z,Y).

per:siblings(X,Y) :- isSibling(Y), relativeOf(X, Y).
per:siblings(X,Y) :- per:parents(Z,X), per:parents(Z,Y), X!=Y.
per:siblings(X,Y) :- partOfMany(X,Z), per:siblings(Z,Y), X!=Y.
Slot Filling Architecture

- **Named mention (PER, ORG, GPE)**
- **DB id (optional)**
- **Document**

**Document Search**

**Relation Inference**

**Slot Candidate Selection**

**SLOTS**

Candidate documents

Candidate slots
Slot Filling: Results

0.794 by returning NIL to all queries
Conclusion

- Combination of statistical and rule-based inference models for entity linking and slot filling
- Systems performed well on slot filling, modestly on entity linking
- There’s no data like more data