



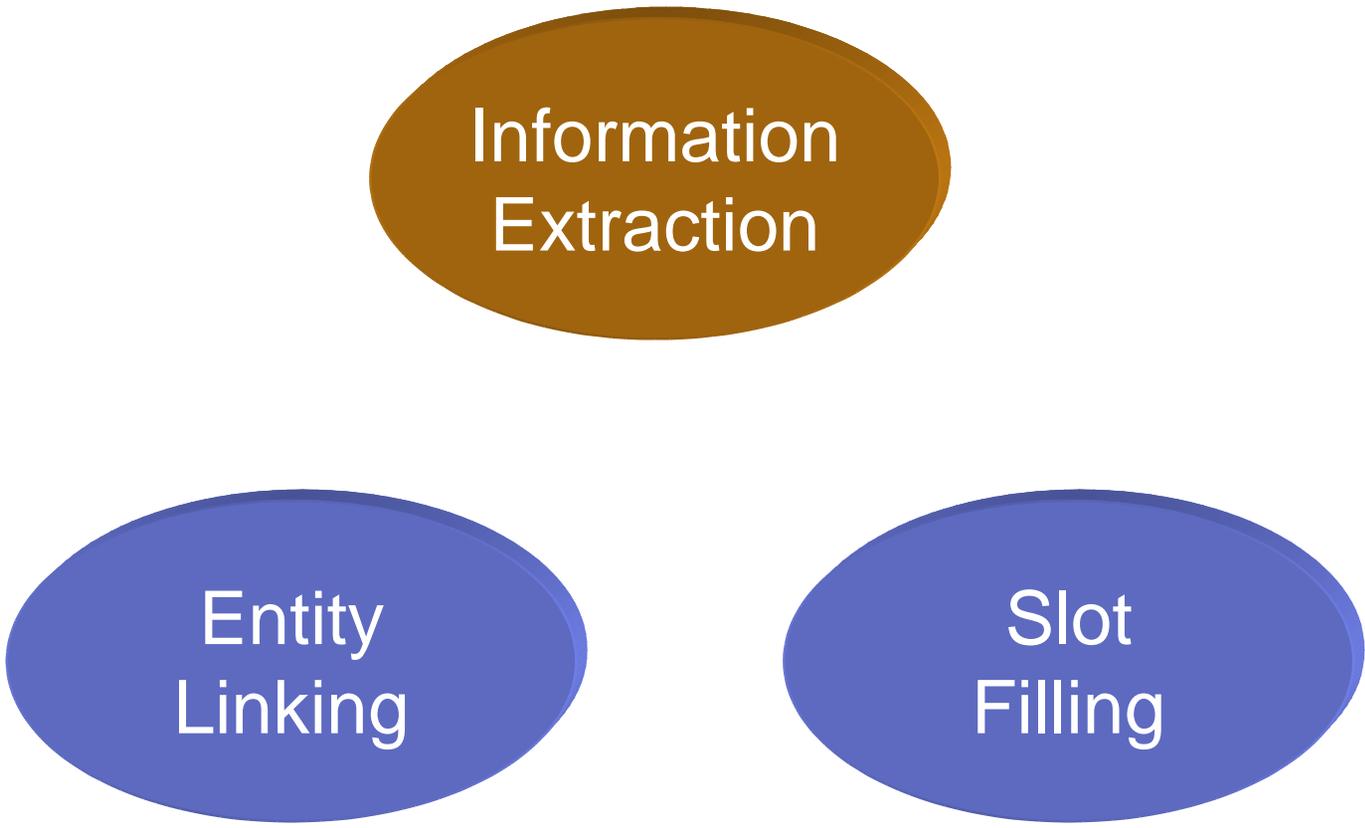
Natural Language Systems Lab

# Entity Linking and Slot Filling through Statistical Processing and Inference Rules

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

November 16<sup>th</sup>, 2009

## Overview



Information  
Extraction

Entity  
Linking

Slot  
Filling

## Information Extraction

<TEXT>

<P>

ABBOTT PARK, Illinois

</P>

<P>

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.

</P>

<P>

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

Tokenization

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

Tokenization

Sentence  
segmentation

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 ( 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 .

Tokenization

Sentence  
segmentation

Case  
restoration

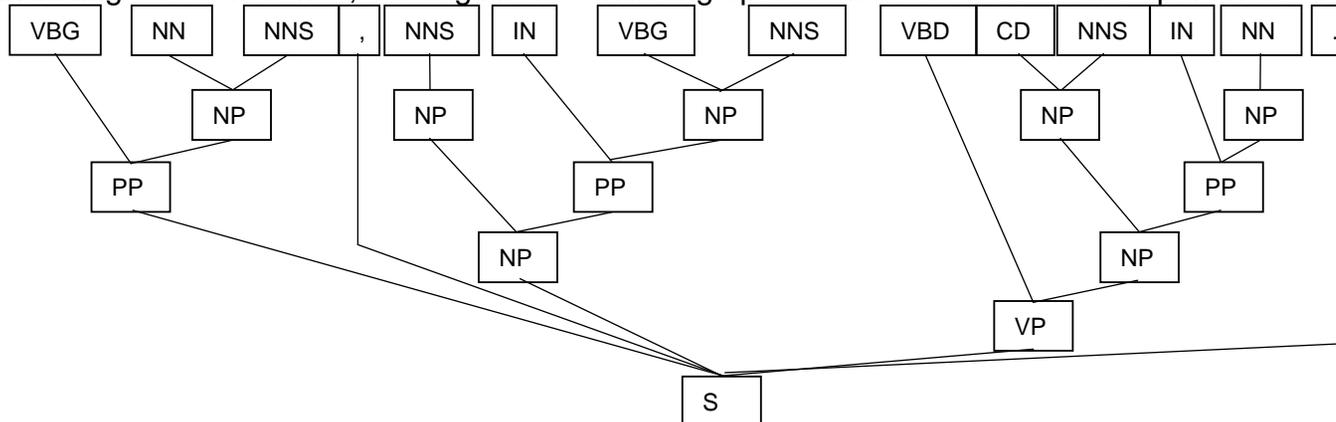
# Information Extraction

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 ( 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 .



- Tokenization
- Sentence segmentation
- Case restoration
- Parsing
- Semantic role labeling

# 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 (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.

Tokenization

Sentence segmentation

Case restoration

Parsing

Semantic role labeling

Mention Detection

# 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 (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.

Tokenization

Sentence segmentation

Case restoration

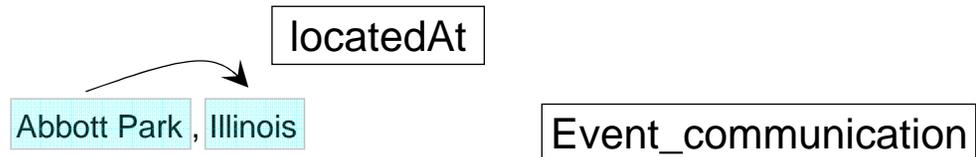
Parsing

Semantic role labeling

Mention Detection

Coreference resolution

# Information Extraction



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 .

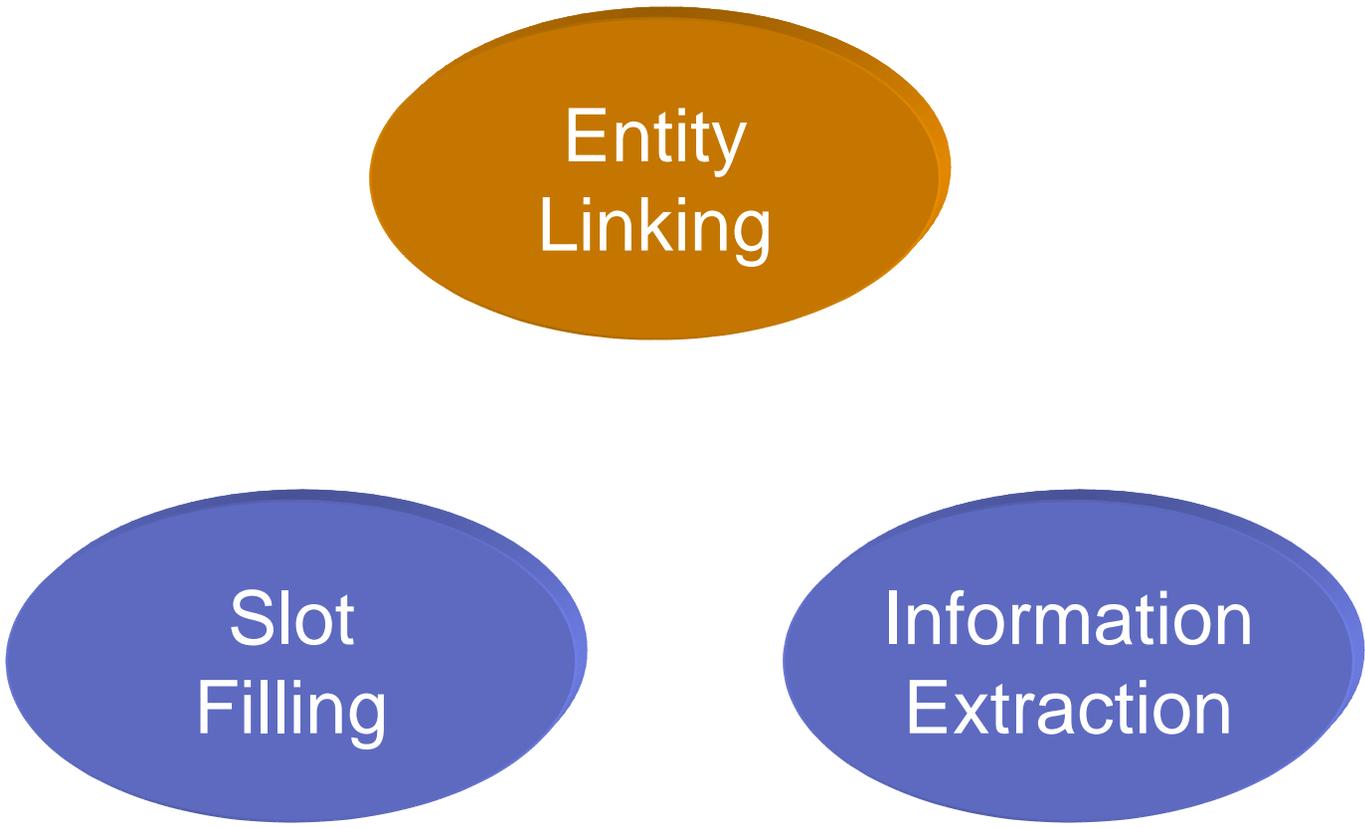
Excluding one-time items , earnings from continuing operations totaled 58 cents per share .

- Tokenization
- Sentence segmentation
- Case restoration
- Parsing
- Semantic role labeling
- Mention Detection
- Coreference resolution
- Relation Detection

## 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

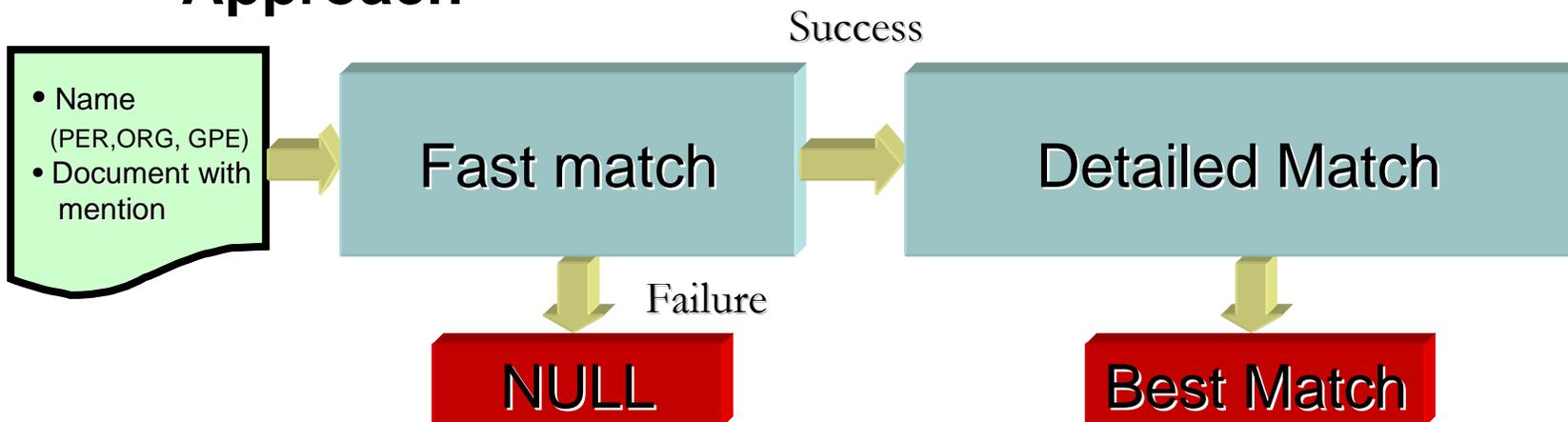
The diagram consists of three ovals. The top oval is orange and contains the text 'Entity Linking'. Below it are two blue ovals, one on the left containing 'Slot Filling' and one on the right containing 'Information Extraction'. The ovals are arranged in a triangular pattern, suggesting a hierarchy or flow of progress.

Slot  
Filling

Information  
Extraction

## Entity Linking: Overall Approach

### ■ Approach



- 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

Type	Count
PER	114,523
ORG	55,813
GPE	116,498
UKN	531,907
<b>Total</b>	<b>818,741</b>

### DBpedia

- v3.2, which contains an ontology

Type	Count
PER	211,029
ORG	75,627
GPE	179,842
<b>Total</b>	<b>466,498</b>

## 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”

## Detailed match for context

### Query Context

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 .

## 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/](http://www.abbott.com/)

## Detailed match for context

### Query Context

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

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

### DB Context

## – Context similarity score:

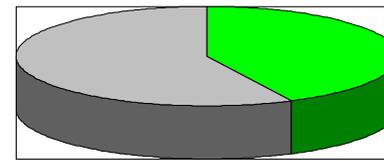
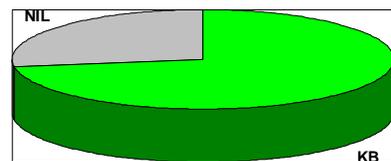
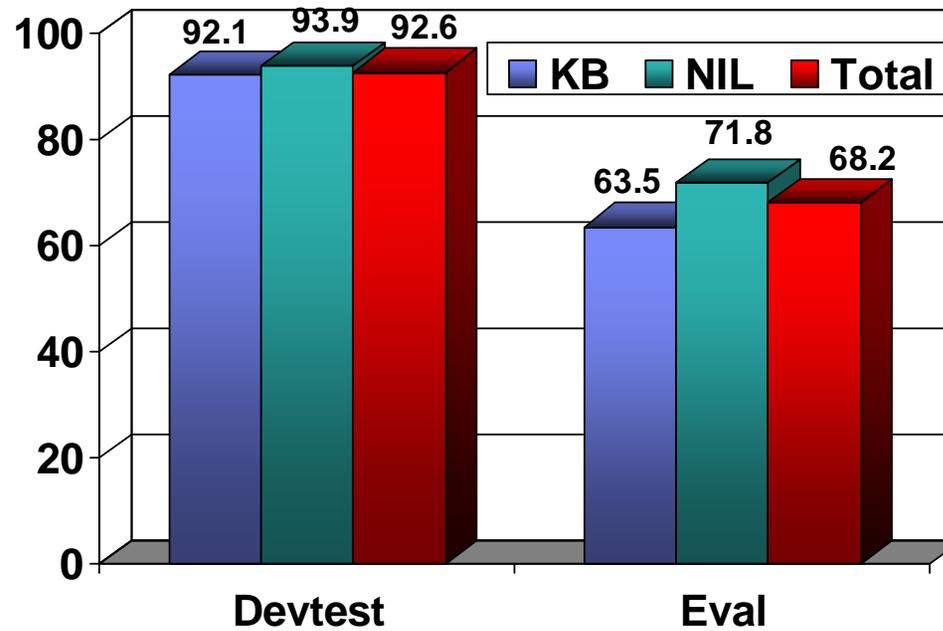
- word inclusion of the query context in the DB context

## Detailed match: overall similarity score

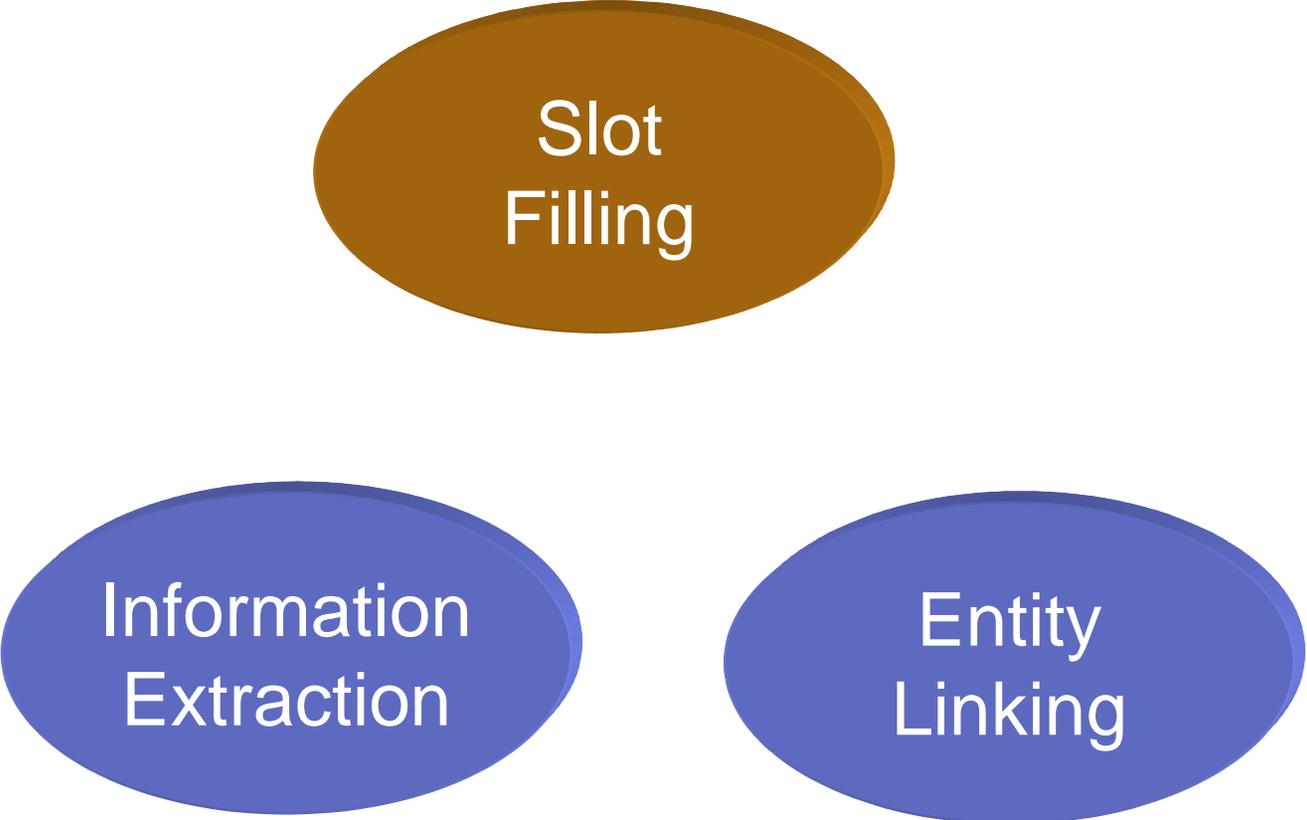
- **Overall similarity score**

$$\text{Sim}(e_1, e_2) = \text{NameSim}(e_1, e_2) \cdot \text{ContextSim}(e_1, e_2)^\alpha$$

# Entity Linking Results



## 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

Person	Organization	Geo-Political Entity
per:alternate_names	org:alternate_names	gpc:alternate_names
per:date_of_birth	org:political/religious_affiliation	gpc:capital
per:age	org:top_members/employees	gpc:subsidiary_orgs
per:place_of_birth	org:number_of_employees/members	gpc:top_employees
per:origin	org:members	gpc:political_parties
per:date_of_death	org:member_of	gpc:established
per:place_of_death	org:subsidiaries	gpc:population
per:cause_of_death	org:parents	gpc:currency
per:residences	org:founded_by	
per:schools_attended	org:founded	
per:title	org:dissolved	
per:member_of	org:headquarters	
per:employee_of	org:shareholders	
per:religion	org:website	
per:spouse		
per:children		
per:parents		
per:siblings		
per:other_family		
per:charges		

## 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).

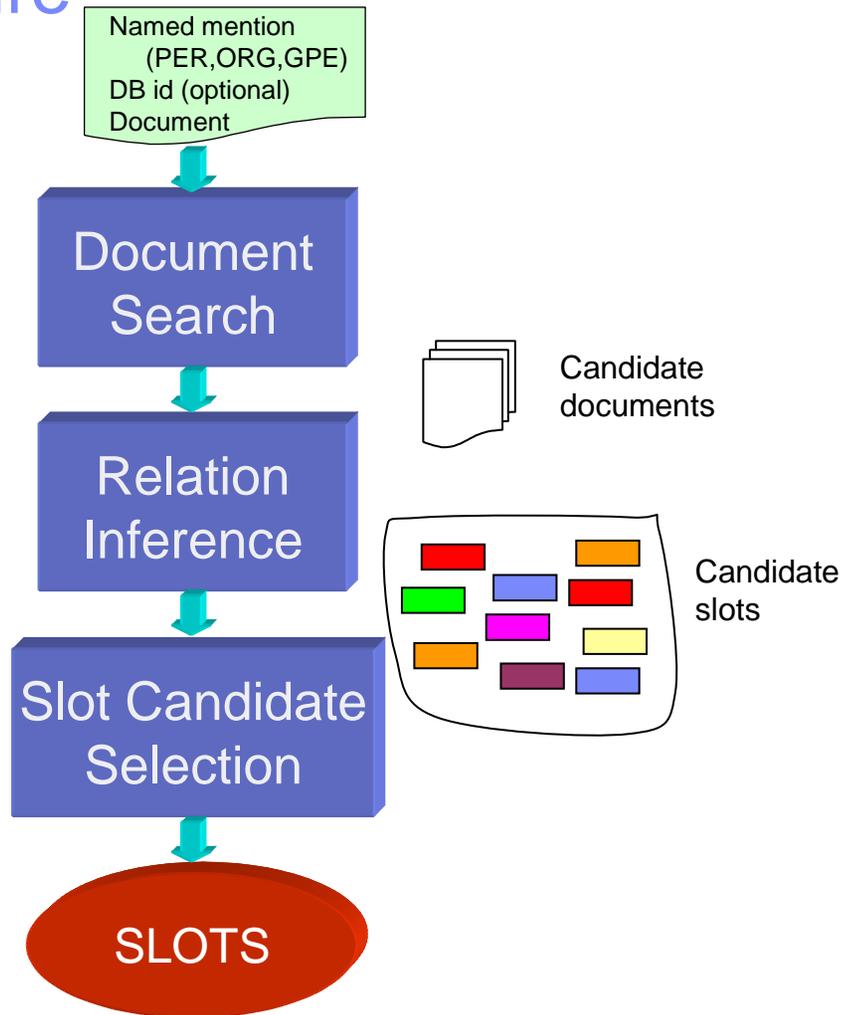
per:religion(X,Y) :- partOfMany(X,Y), religious(Y).  
per:religion(X,Y) :- located(X,Z), religiousFacility(Z,Y).

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.

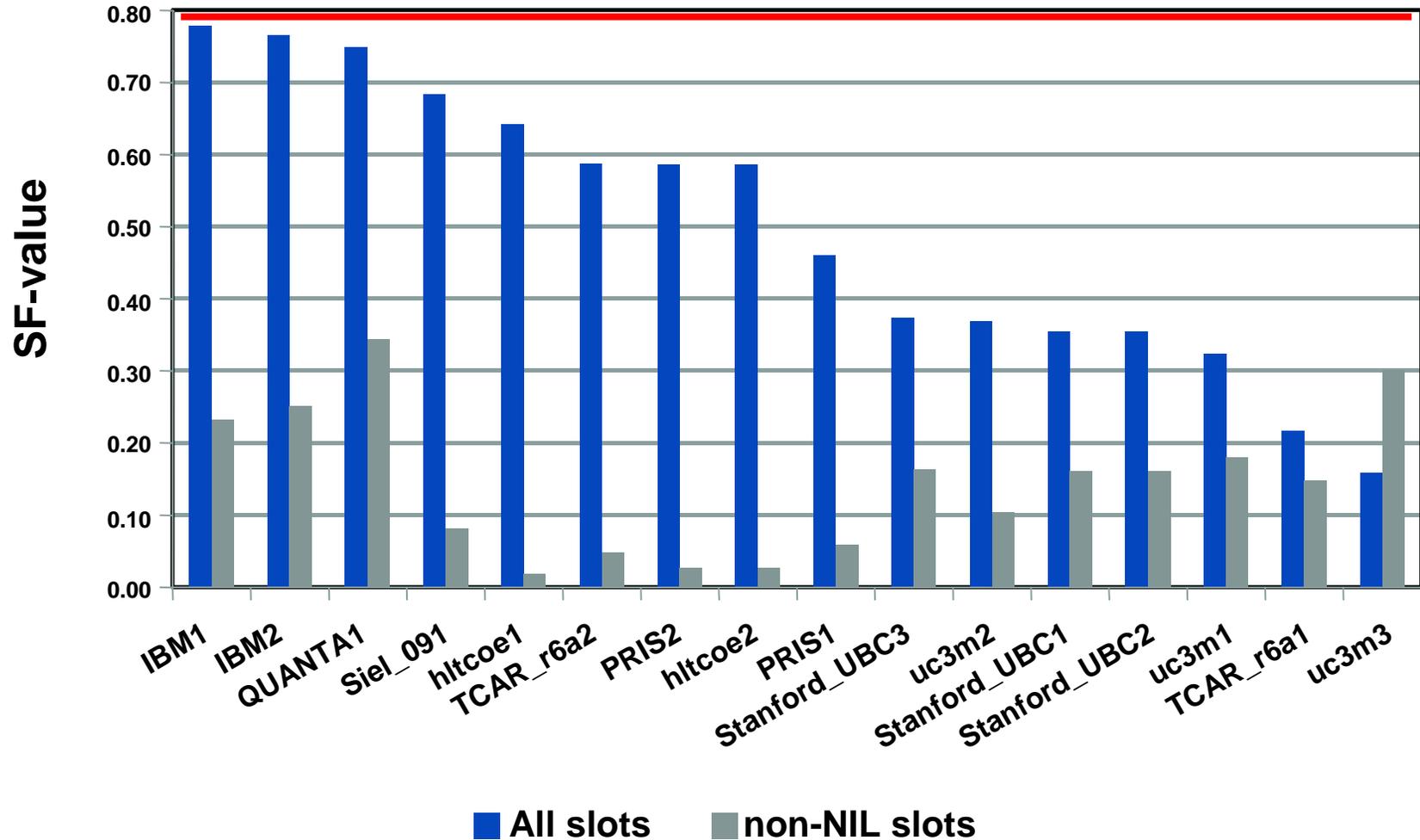
# 130 rules

# Slot Filling Architecture



# 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**