DEANNA: Natural Language Questions for the Web of Data

Mohamed Yahya†
Maya Ramanath‡
† Max Planck Institute for Informatics

Klaus Berberich†
Volker Tresp#
*Qatar Computing Research Institute

Shady Elbassiousni*
Gerhard Weikum†
‡IIT Delhi
# Siemens
Ingrid Bergman (29 August 1915 – 29 August 1982) was a Swedish actress who starred in a variety of European and American films. She won three Academy Awards, and the Tony Award for Best Actress, making her the fourth greatest female star of American cinema ranked by the American Film Institute. She is best remembered for her roles as Ilsa Lund in Casablanca (1942), an Academy Award-winning World War II drama co-starring Humphrey Bogart, and as Alicia Huberman in Notorious (1946), an Alfred Hitchcock thriller co-starring Cary Grant. In 1950, after a decade of stardom in American films, she starred in the Italian film Stromboli, which led to a love affair with director Roberto Rossellini while they were both already married. The affair and then marriage with Rossellini created a scandal that forced her to remain in Europe until 1956, when she made a successful Hollywood return in Anastasia, for which she won her second Academy Award, as well as the forgiveness of her fans. Many of her personal and film documents can be seen in the Wesleyan University Cinema Archives.

Roberto Gastone Zeffiro Rossellini (8 May 1906 – 3 June 1977) was an Italian film director and screenwriter. Rossellini was one of the directors of the Italian neorealist cinema, contributing films such as Roma città aperta (Rome, Open City 1945) to the movement. His mother, Elettra (née Bellan), was a housewife, and his father, Angiolo Giuseppe "Beppino" Rossellini, owned a construction firm. His mother was of part French descent, from immigrants who had arrived in Italy during the Napoleonic Wars. He lived on the Via Ludovisi, where Benito Mussolini had his first Roman hotel in 1922 when Fascism obtained power in Italy.
WordNet

LOD: > 31 BILLION triples

YAGO2: 120 million facts on 10 million entities

Subject
Rome
city
Roberto_Rossellini

Predicate
isA
subclassOf
marriedTo

Object
city
location
Ingrid_Bergman
What is Toronto???
Who played in *Casablanca* and was married to a writer born in Rome?

```
SELECT ?p WHERE{
  ?p type person .
  ?p actedIn Casablanca_(film) .
  ?p isMarriedTo ?w .
  ?w type writer .
  ?w bornIn Rome }
```

- Dot: conjunction
- ?p, ?w: variables to be bound
- Same variable → same binding (join)
What is DEANNA?

Question: “Who played in Casablanca and was married to a writer born in Rome?”

DEANNA

SPARQL

KB

Answers

?p type person.
?p actedIn Casablanca_(film).
?p isMarriedTo ?w.
?w type writer.
?w bornIn Rome.
Inside DEANNA

DEANNA

Question

DEANNA

SPARQL

KB

Answers

Phrase detection

Phrase mapping

Dependency detection

Joint Disambig.

Query Generation

Phrase detection

Phrase mapping

Dependency detection

Query Generation

Answer
Inside DEANNA

Question

DEANNA

SPARQL

KB

Answers

Phrase detection

Phrase mapping

Dependency detection

Joint Disambig.

Query Generation

Phrase detection

Phrase mapping

Dependency detection

Joint Disambig.

Query Generation

DEANNA

SPARQL

KB

Answers

q1

Rome

was born

r:bornInPlace

c:writer

Casablanca

played in

r:actedIn

c:person

Who

was married to

r:isMarriedTo

q2

q3
Structured Query Generation

SELECT ?p WHERE {
  ?p type person.
  ?p actedIn Casablanca_(film).
  ?p isMarriedTo ?w.
  ?w type writer.
  ?w bornIn Rome
}
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Disambiguation graph construction
Disambiguation graph processing

Question
DEANNA
SPARQL
KB

Experiments & Results
**Outline**

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

DEANNA

SPARQL

KB

Experiments & Results
Phrase Detection

**Concepts:** entities & classes: Dictionary-based

<table>
<thead>
<tr>
<th>Concept</th>
<th>Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casablanca</td>
<td>Casablanca</td>
</tr>
<tr>
<td>Casablanca</td>
<td>Casablanca, Morocco</td>
</tr>
<tr>
<td>Casablanca_(film)</td>
<td>Casablanca the film</td>
</tr>
<tr>
<td>Casablanca_(film)</td>
<td>Casablanca</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Relations:**

Mainly use Reverb [Fader et al. EMNLP’11]: \( V \mid VP \mid VWP \)

... \textit{was/VBD married/VBN to/TO} a/DT...
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Experiments & Results

Question

DEANNA

SPARQL

KB
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

DEANNA

SPARQL

KB

Experiments & Results
Phrase Mapping

• **Concepts**: entities & classes: Dictionary-based

<table>
<thead>
<tr>
<th>Concept</th>
<th>Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casablanca</td>
<td>Casablanca</td>
</tr>
<tr>
<td>Casablanca</td>
<td>Casablanca, Morocco</td>
</tr>
<tr>
<td>Casablanca_(film)</td>
<td>Casablanca the film</td>
</tr>
<tr>
<td>Casablanca_(film)</td>
<td>Casablanca</td>
</tr>
<tr>
<td>Played_(film)</td>
<td>Played</td>
</tr>
</tbody>
</table>

• **Relations**: Dictionary-based

<table>
<thead>
<tr>
<th>Relation</th>
<th>Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>actedIn</td>
<td>acted in</td>
</tr>
<tr>
<td>actedIn</td>
<td>played in</td>
</tr>
<tr>
<td>hasMusicalRole</td>
<td>plays</td>
</tr>
<tr>
<td>hasMusicalRole</td>
<td>mastered</td>
</tr>
</tbody>
</table>
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Experiments & Results
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Experiments & Results
Look for specific patterns in dependency parses [de Marneffe et al. LREC’06]
RESULT
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Experiments & Results
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Experiments & Results

DEANNA

SPARQL

KB
Joint Disambiguation - ILP

- **ILP:** Integer Linear Programming

- maximize $\alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} v_{k,l} Z_{k,l} + \ldots$

- Subject to:
  - No token in multiple phrases,
  - Triples observe type constraints, ...

---

EMNLP
July 12, 2012

Natural Language Questions for the Web of Data - Yahya et al. 23
Joint Disambiguation – Objective

\[ \alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} v_{k,l} Z_{k,l} \]

Prior
Joint Disambiguation – Objective

\[ \alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} v_{k,l} Z_{k,l} \]

Semiotic nodes

Phrase nodes

q-nodes

q\_1

born

was born

a writer

Similarity Edges

Semantic nodes

e:Rome

e:Sydne_Rome

e:Born_(film)

e:Max_Born

r:bornOnDate

r:bornInPlace

c:writer

Coherence Edges

Coherence
A phrase node can be assigned to one semantic node:

\[ \alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} \nu_{k,l} Z_{k,l} \]
Joint Disambiguation – Constraints

Classes translate to type-constrained variables

→ Every semantic triple should have a class to \textit{join} & \textit{project}!

\text{person} \text{ actedIn } \text{Casablanca\_\text{(film)}}

\text{?x type person} \quad . \quad \text{?x actedIn Casablanca\_\text{(film)}}
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Question

DEANNA

SPARQL

KB

Experiments & Results
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Experiments & Results

DEANNA

SPARQL

KB

Query Generation

Question
Structured Query Generation

```
SELECT ?p WHERE {
  ?w type writer .
  ?w bornIn Rome .
  ?p type person .
  ?p actedIn Casablanca_(film) .
  ?p isMarriedTo ?w }
```
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

Question

DEANNA

SPARQL

KB

Experiments & Results
Outline

1. Phrase detection
2. Phrase mapping
3. Dependency detection
5. Query Generation

DEANNA

SPARQL

KB

Experiments & Results
Evaluation Methodology

• **3-stage** evaluation for more insight:
  1. Disambiguation
  2. Query generation
  3. Query answering

• We rely on **human judges**
Datasets

• QALD-1
  ▪ YAGO2
  ▪ 27/50 questions within scope
  ▪ “Which software has been published by Mean Hamster Software?”

• NAGA [Elbassuoni et al. CIKM’09]
  ▪ YAGO+IMDB
  ▪ 44/87 questions within scope
  ▪ “Which director has won the Academy Award for Best director and is married to an actress that has won the Academy Award for Best Actress?”
Results 1/3: Disambiguation

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>QALD-1</th>
<th>NAGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$cov_{\text{macro}}$</td>
<td>0.973</td>
<td>0.934</td>
</tr>
<tr>
<td>$prec_{\text{macro}}$</td>
<td>1.000</td>
<td>0.934</td>
</tr>
<tr>
<td>$cov_{\text{micro}}$</td>
<td>0.963</td>
<td>0.945</td>
</tr>
<tr>
<td>$prec_{\text{micro}}$</td>
<td>1.000</td>
<td>0.941</td>
</tr>
</tbody>
</table>

$cov = \text{correct/ideal}$

$prec = \text{correct / retrieved}$
Results 2/3: Query Generation

**Issues:**
- Incorrect disambiguation
- Incorrect dependencies.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>QALD-1</th>
<th>NAGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{cov}_{\text{macro}}$</td>
<td>0.975</td>
<td>0.894</td>
</tr>
<tr>
<td>$\text{prec}_{\text{macro}}$</td>
<td>1.000</td>
<td>0.941</td>
</tr>
<tr>
<td>$\text{cov}_{\text{micro}}$</td>
<td>0.963</td>
<td>0.847</td>
</tr>
<tr>
<td>$\text{prec}_{\text{micro}}$</td>
<td>1.000</td>
<td>0.906</td>
</tr>
</tbody>
</table>

$cov = \frac{\text{#correct triples}}{\text{#ideal triples}}$

$prec = \frac{\text{#correct triples}}{\text{#retrieved triples}}$
Results 3/3: Answering

**Relaxation:** Keep type constraints, everything else to keywords

\(?x\) type writer . \(?x\) bornIn Rome

\(?x\) type writer . \(?x\) bornIn \(?y\) [“Rome”]

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>QALD-1</th>
<th>NAGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>#questions</td>
<td>27</td>
<td>44</td>
</tr>
<tr>
<td>#queries</td>
<td>20</td>
<td>41</td>
</tr>
<tr>
<td>#satisfactory</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>#relaxed</td>
<td>+3</td>
<td>+3</td>
</tr>
</tbody>
</table>

**unsatisfactory:** main problem is *empty result set*
“Who played in Casablanca and was married to a writer born in Rome?”

```
?p type person.
?p actedIn Casablanca_(film).
?p isMarriedTo ?w.
?w type writer.
?w bornIn Rome.
```
Thank you.

QUESTIONS/COMMENTS?

bit.ly/mpi-deanna