

DEANNA: Natural Language Questions for the Web of Data

Mohamed Yahya[†]

Maya Ramanath[‡]

Klaus Berberich[†]

Volker Tresp[#]

Shady Elbassiousni^{*}

Gerhard Weikum[†]

[†] Max Planck Institute
for Informatics

^{*} Qatar Computing
Research Institute

[‡] IIT Delhi

[#] Siemens



“Classical” QA

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

“Classical” QA

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

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, two Emmy Awards, and the Tony Award for Best Actress. She is ranked as the fourth greatest female star of American cinema of all time by the American Film Institute. She is best remembered for her **roles as Ilsa Lund in *Casablanca* (1942)**, a World War II drama co-starring Humphrey Bogart and as Alicia Huberman in *Notorious* (1946), an Alfred Hitchcock thriller co-starring Cary Grant.

“Classical” QA

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

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, two Emmy Awards, and the Tony Award for Best Actress. She is ranked as the fourth greatest female star of American cinema of all time by the American Film Institute. She is best remembered for her **roles as Ilsa Lund in *Casablanca*** (1942), a World War

II drama
in *Notre-Dame de Paris*
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. [\[4\]](#)

“Classical” QA

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

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, two Emmy Awards, and the Tony Award for Best Actress. She is ranked as the fourth greatest female star of American cinema of all time by the American Film Institute. She is best remembered for her **roles as Ilsa Lund in *Casablanca*** (1942), a World War

II dram
in *Not
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 r
Academy Av
personal and
University C

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.

“Classical” QA

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

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. She is ranked the fourth greatest female star of American cinema by the American Film Institute. She is best remembered for her roles as **Ilsa Lund in *Casablanca* (1942)**, and as **Irma in *Notorious* (1946)**.

II dram
in *Noto*
Grant.

In 1950, after a decade of stardom in Hollywood, she married the Italian film *Stromboli*, which led to her divorce from 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 return to Hollywood and won her fourth Academy Award for *Annie Hall*. Her personal and professional life was the subject of a 1972 book by University of California Press.

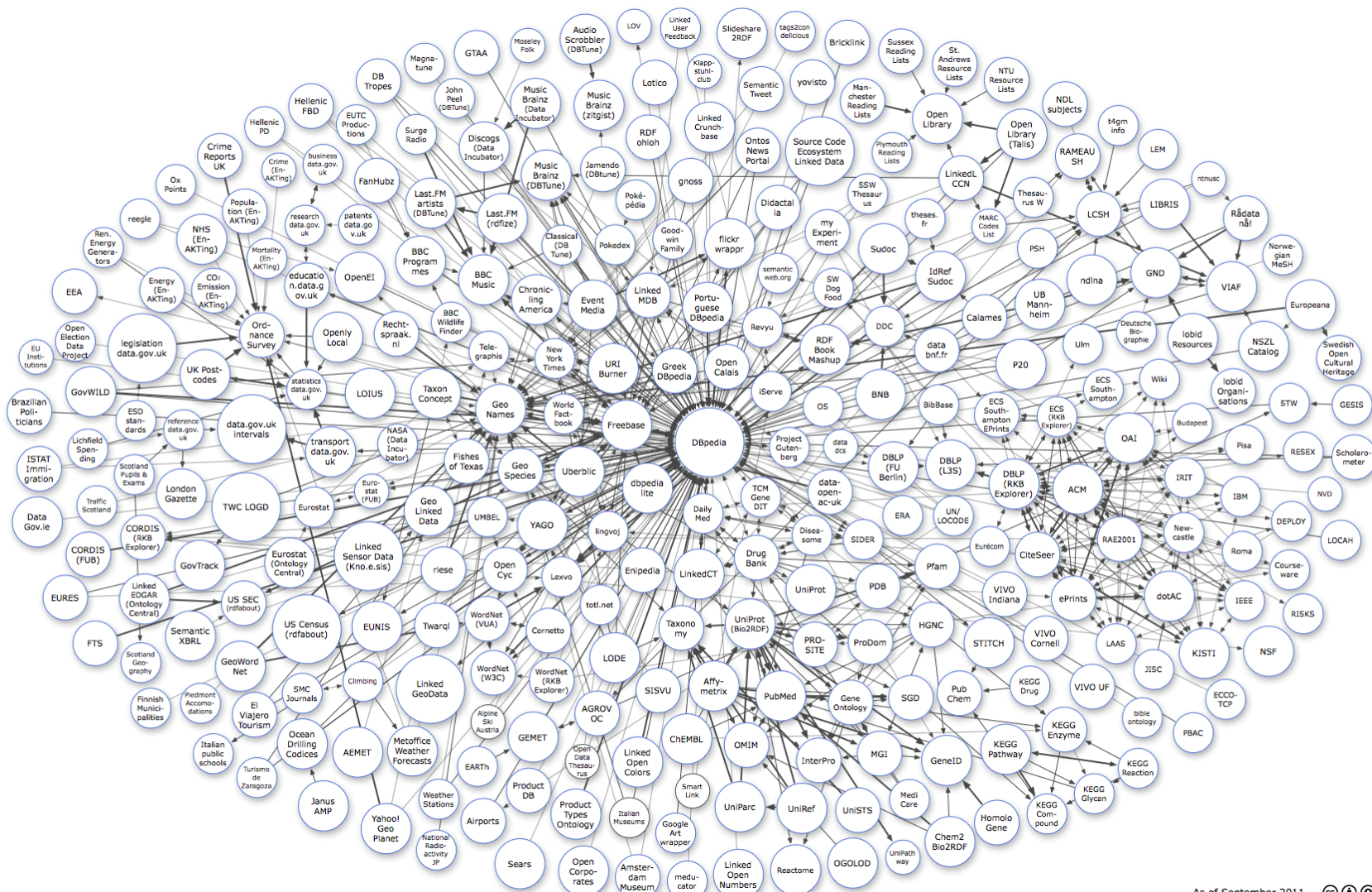
director Roberto Rossellini while they were both already married. The affair and then **marriage with Rossellini** created a scandal that forced her to remain in Europe until 1956, when she made a successful return to Hollywood and won her fourth Academy Award for *Annie Hall*. Her personal and professional life was the subject of a 1972 book by University of California Press.

Hollywood r
Academy Av
personal an
University C

Rossellini was born in Rome. 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.

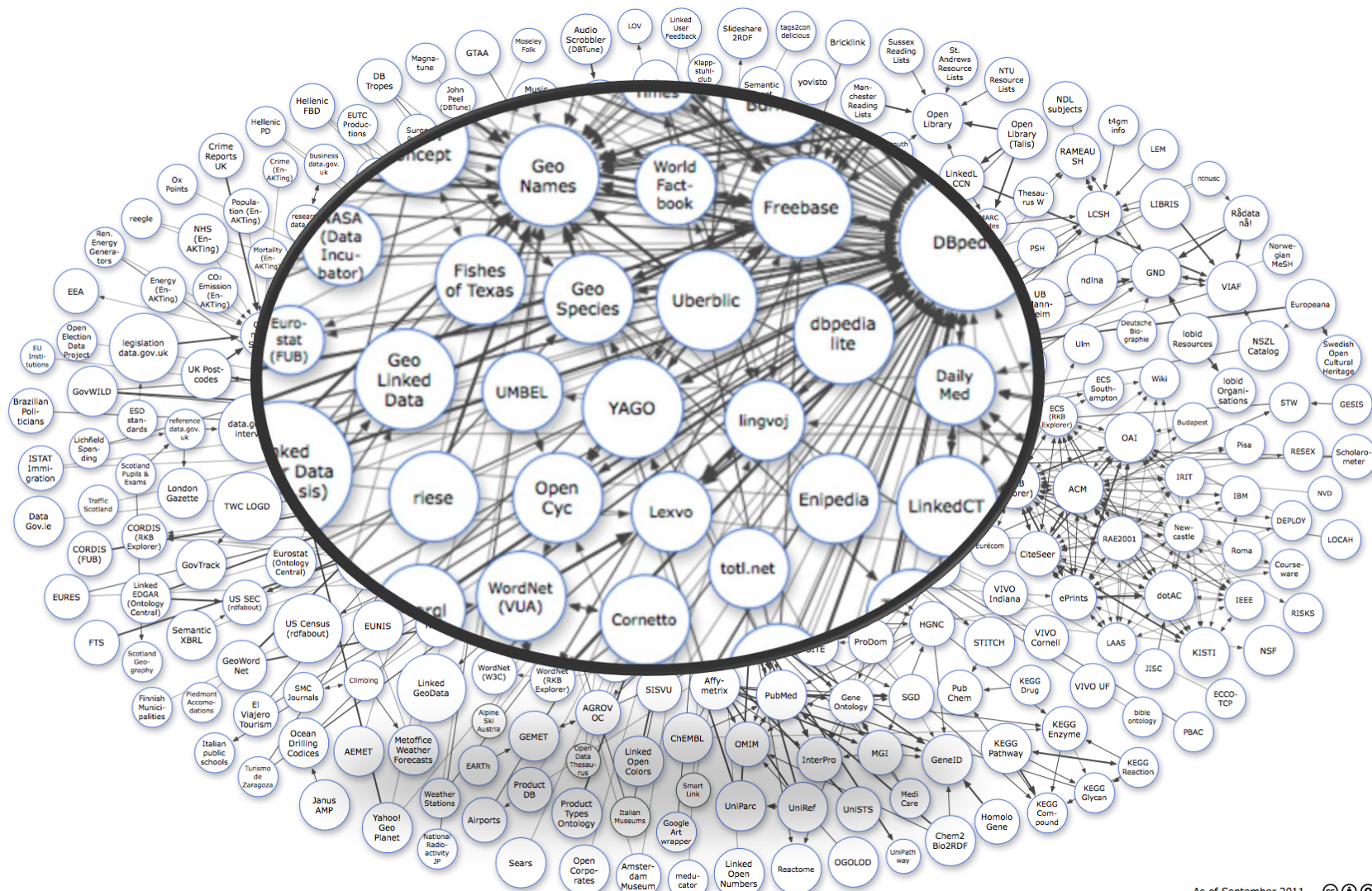
Roberto Gastone Zeffiro Rossellini (8 May 1906 – 3 June 1977) was an Italian film director and **screenwriter**. Rossellini was one of the directors of the Italian neorealist cinema, contributing films such as *Roma città aperta* (*Rome, Open City* 1945) to the movement.

QA, meet the (semantic) Web of Data



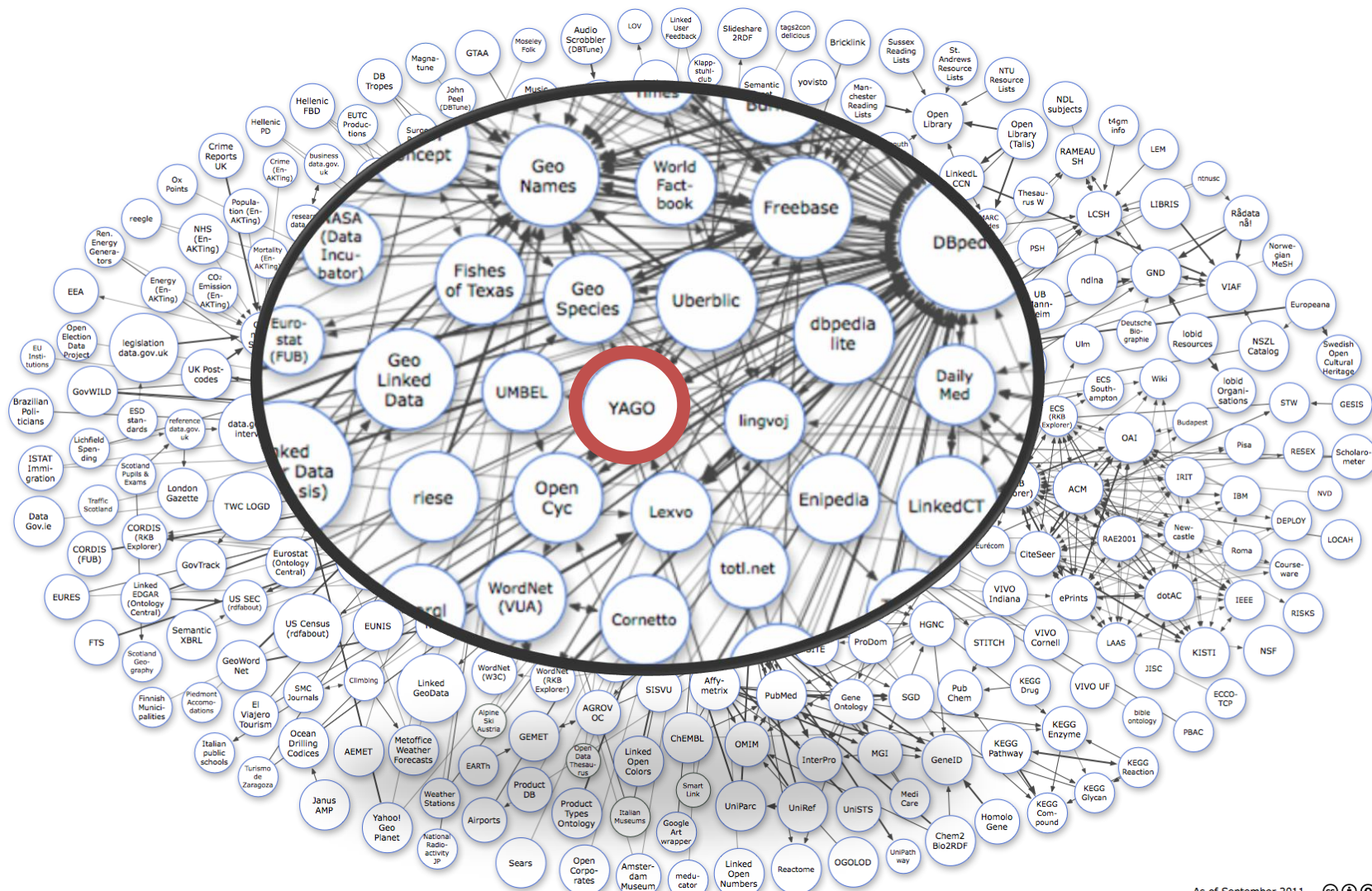
As of September 2011

QA, meet the (semantic) Web of Data



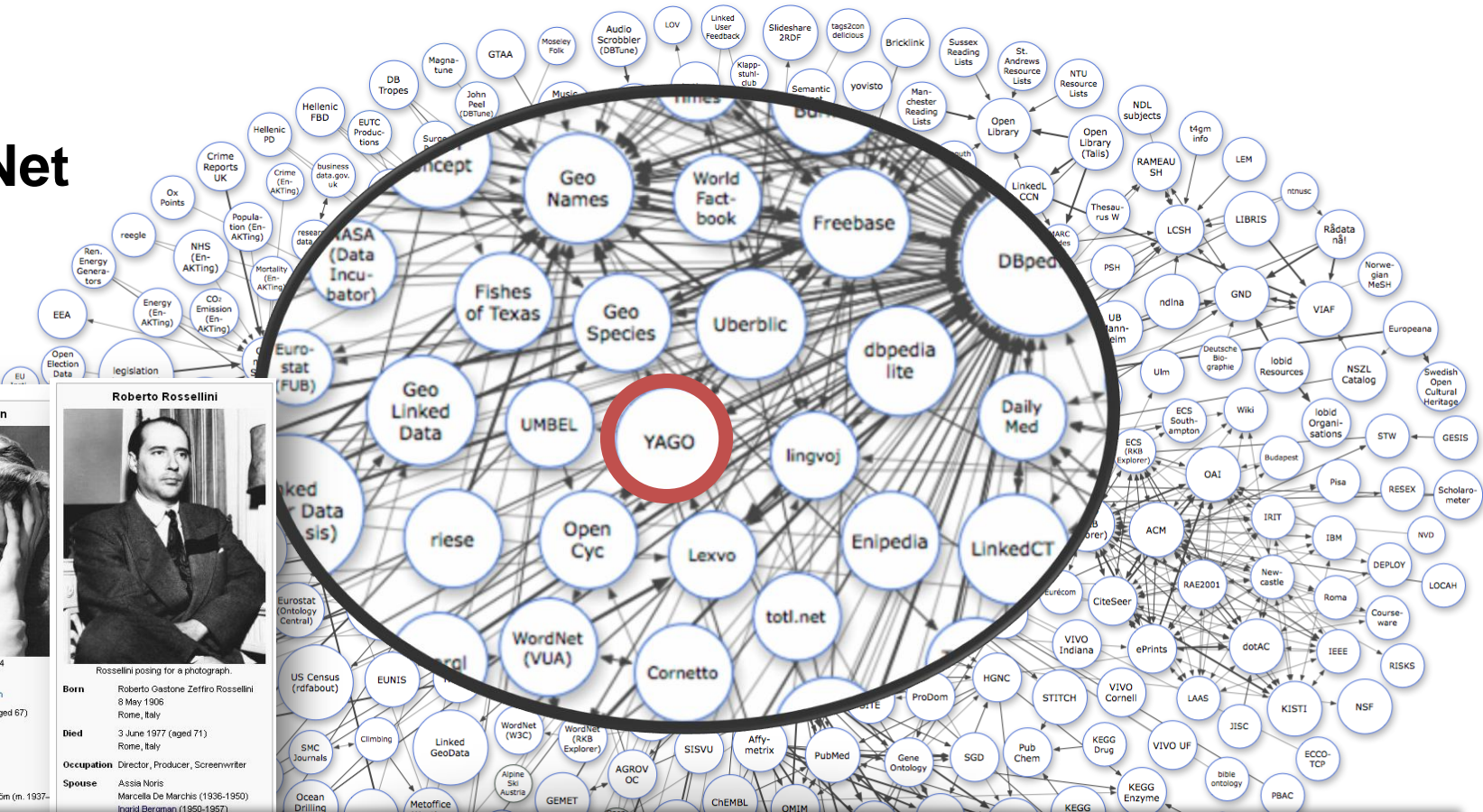
As of September 2011

QA, meet the (semantic) Web of Data



As of September 2011

WordNet



Rome

city

Roberto_Rosellini

Predicate

isa

subclassof

marriedTo

Object

city

location

Ingrid_Bergman

WordNet



Roberto_Rosellini

marriedTo

Ingrid_Bergman

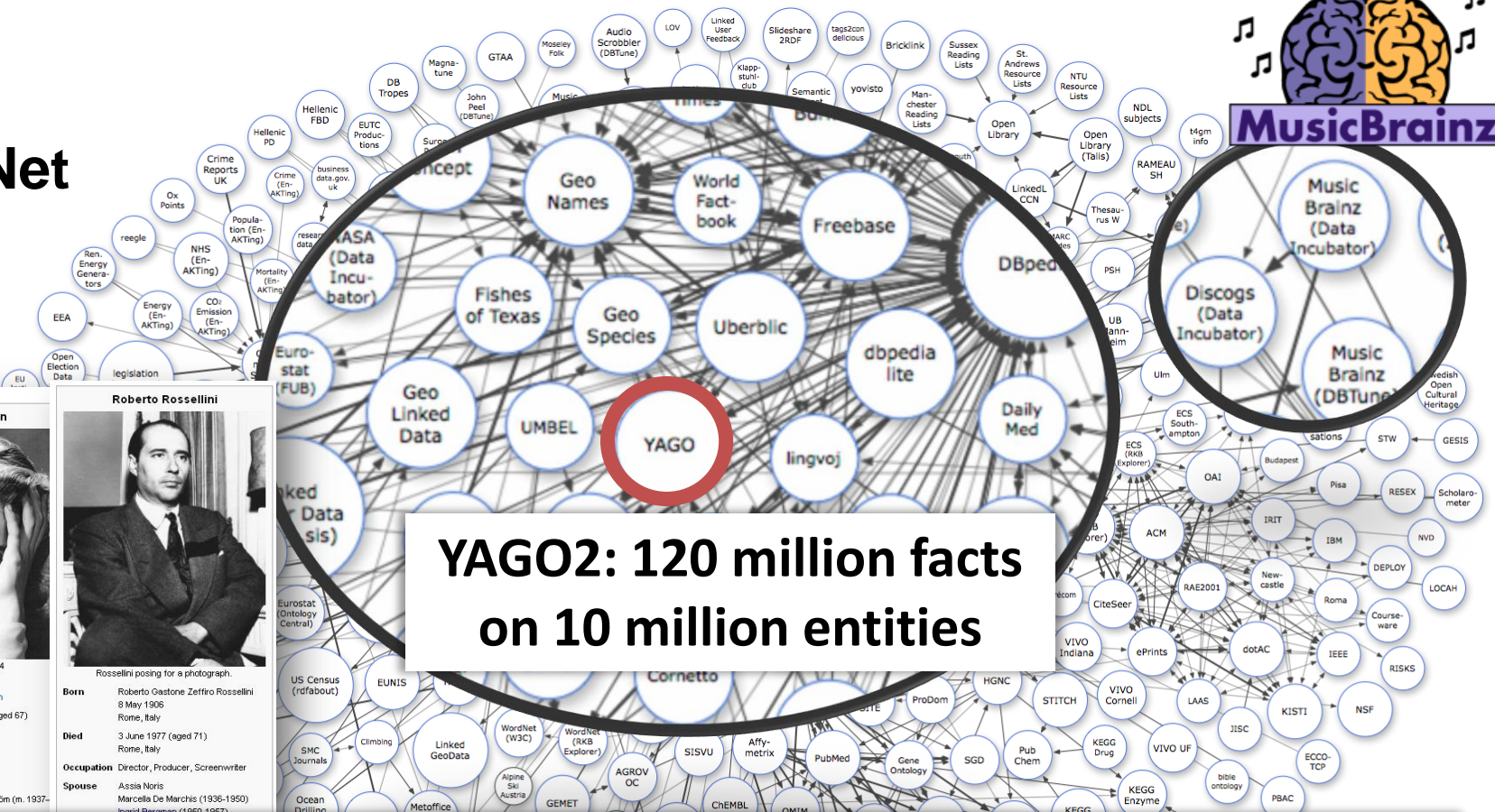
WordNet



	Ingrid Bergman, 1944
Born	29 August 1915 Stockholm, Sweden
Died	29 August 1982 (aged 67) London, England
Occupation	Actress
Years active	1934–82
Spouse	Petter Aron Lindström (m. 1937 (div.))



Rossellini posing for a photograph.	
Born	Roberto Gastone Zeffiro Rossellini 8 May 1906 Rome, Italy
Died	3 June 1977 (aged 71) Rome, Italy
Occupation	Director, Producer, Screenwriter
Spouse	Assia Noris Marcella De Marchis (1936-1950) Isabella Rossellini (1959-1963)



Rome

city

Roberto_Rosellini

Predicate

isa

subclassof

marriedTo

Object

city

location

Ingrid_Bergman

QA, meet the (semantic) Web of Data

WordNet

+



Ingrid Bergman	
	
Ingrid Bergman, 1944	
Born	29 August 1915 Stockholm, Sweden
Died	29 August 1982 (aged 67) London, England
Occupation	Actress
Years active	1934–62
Spouse	Petter Aron Lindström (m. 1937–1950)

Roberto Rossellini	
	
Rossellini posing for a photograph.	
Born	Roberto Gastone Zeffiro Rossellini 8 May 1906 Rome, Italy
Died	3 June 1977 (aged 71) Rome, Italy
Occupation	Director, Producer, Screenwriter
Spouse	Assia Noris Marcella De Marchis (1936–1950) Ingrid Bergman (1950–1957)

**YAGO2: 120 million facts
on 10 million entities**

Subject

Rome

city

Roberto_Rossellini

Predicate

isa

subclassof

marriedTo

Object

city

location

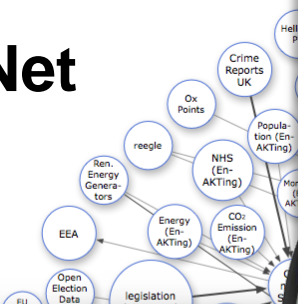
Ingrid_Bergman



MusicBrainz

IMDb

+



Ingrid Bergman

	Ingrid Bergman, 1944
Born	29 August 1915 Stockholm, Sweden
Died	29 August 1982 (aged 67) London, England
Occupation	Actress
Years active	1934–82
Spouse	Petter Aron Lindström (m. 1937– 1950)



Roberto Rossellini

Rossellini posing for a photograph.	
Born	Roberto Gastone Zeffiro Rossellini 8 May 1906 Rome, Italy
Died	3 June 1977 (aged 71) Rome, Italy
Occupation	Director, Producer, Screenwriter
Spouse	Assia Noris Marcella De Marchis (1936-1950) Isabella Rossellini (1959-1967)

LOD: > 31 BILLION triples

**YAGO2: 120 million facts
on 10 million entities**

Predicate



isa

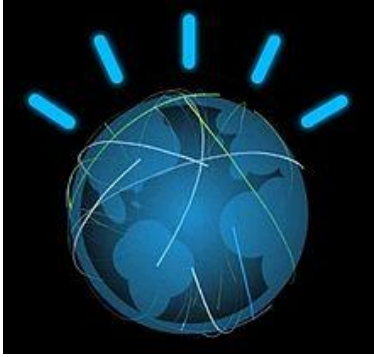
city

subclassof

location

Ingrid_Bergman

Semantics Types



U.S. Cities

*ITS LARGEST AIRPORT WAS
NAMED FOR A WORLD
WAR II HERO; ITS SECOND
LARGEST, FOR A WORLD
WAR II BATTLE*

*What is
Toronto???*

Crash Course: Querying the Web of Data

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

Crash Course: Querying the Web of Data

“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 }
```

SPARQL

Crash Course: Querying the Web of Data

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

Projection Variable

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

SPARQL

Crash Course: Querying the Web of Data

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

Projection Variable

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

SPARQL

Dot: conjunction

Crash Course: Querying the Web of Data

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

Projection Variable

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

SPARQL

Dot: conjunction

?p, ?w: variables to be bound

Crash Course: Querying the Web of Data

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

Projection Variable

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

SPARQL

Dot: conjunction

?p, ?w: variables to be bound

Same variable → same binding (join)

Crash Course: Querying the Web of Data

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

Projection Variable

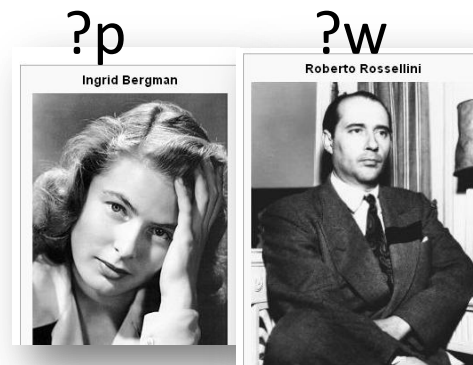
```
SELECT ?p WHERE{  
  ?p type person .  
  ?p actedIn Casablanca_(film) .  
  ?p isMarriedTo ?w .  
  ?w type writer .  
  ?w bornIn Rome }
```

SPARQL

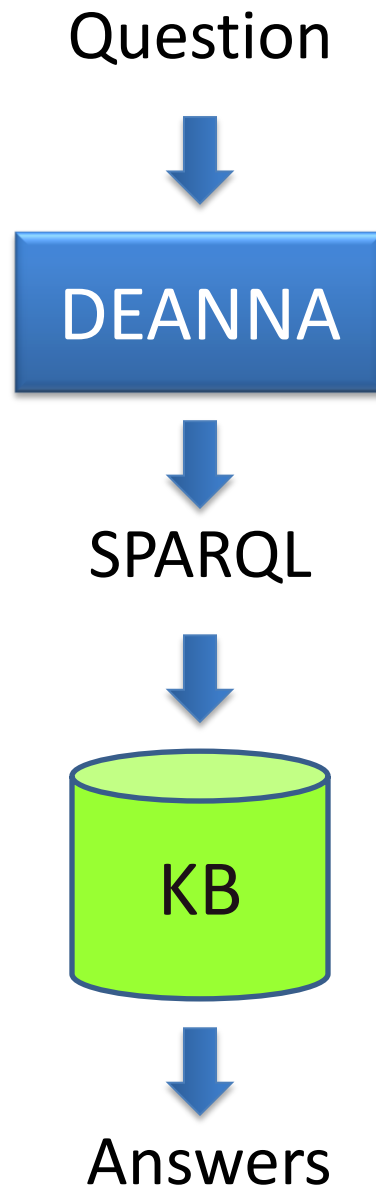
Dot: conjunction

?p, ?w: variables to be bound

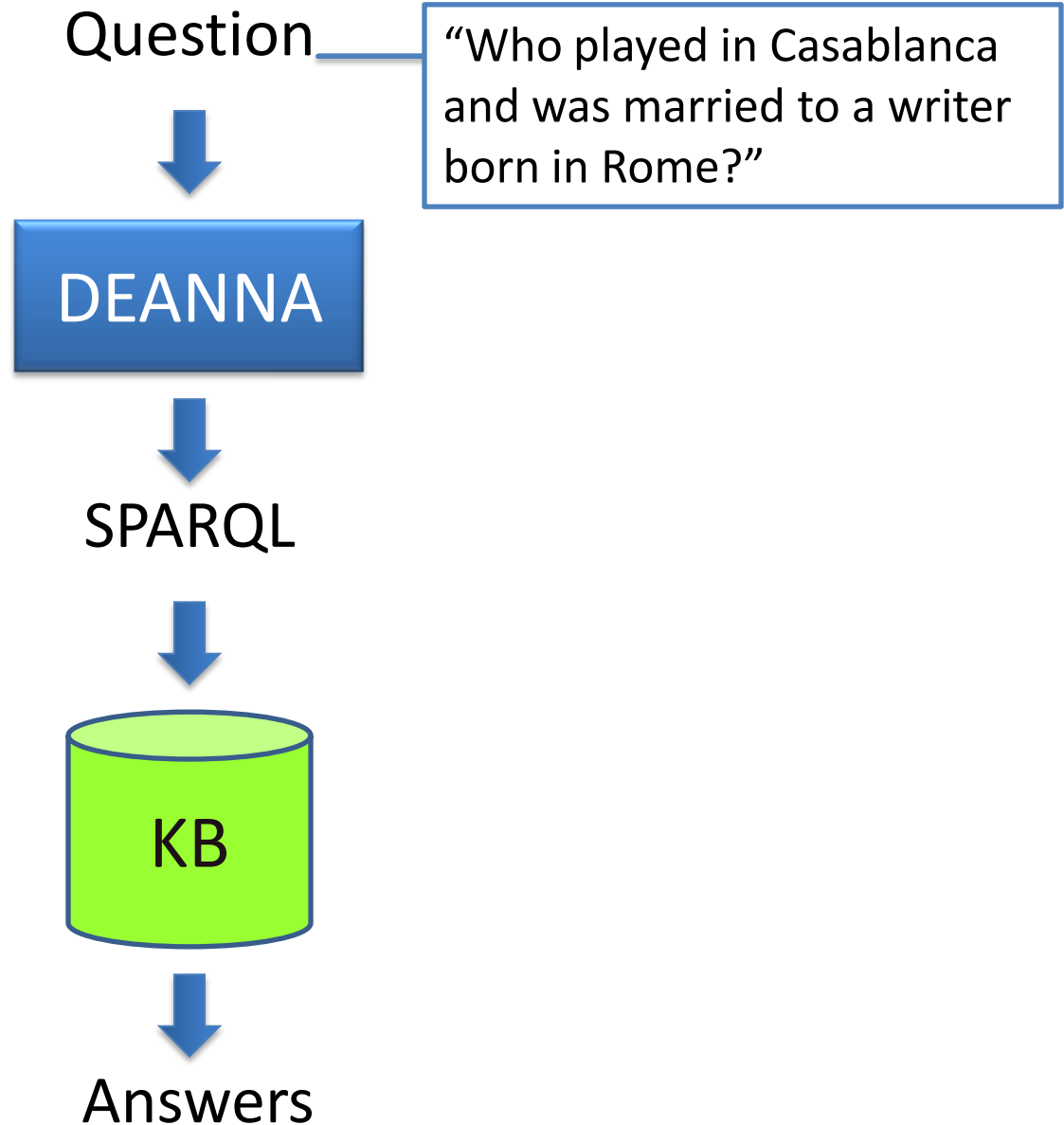
Same variable → same binding (join)



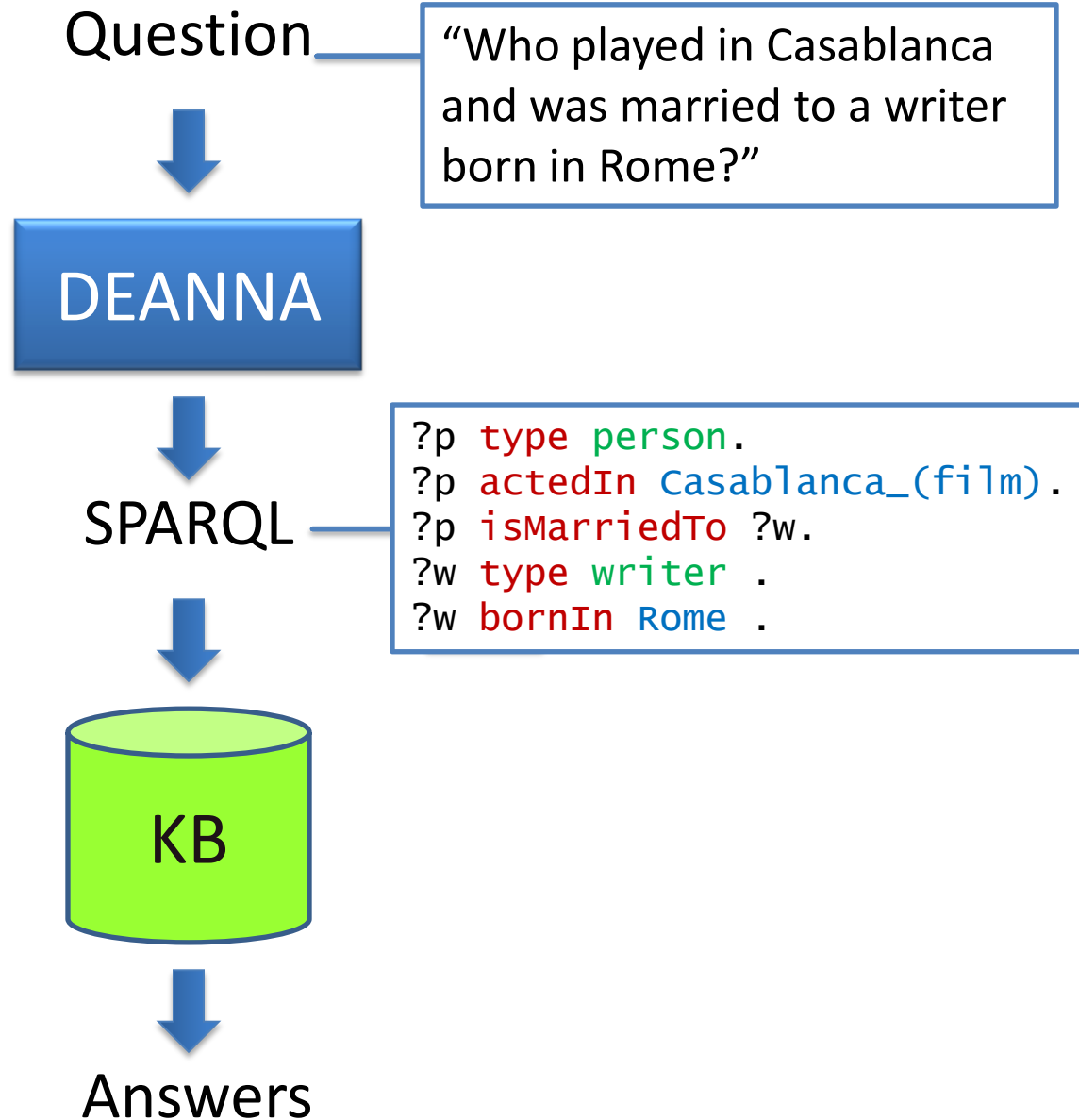
What is DEANNA?



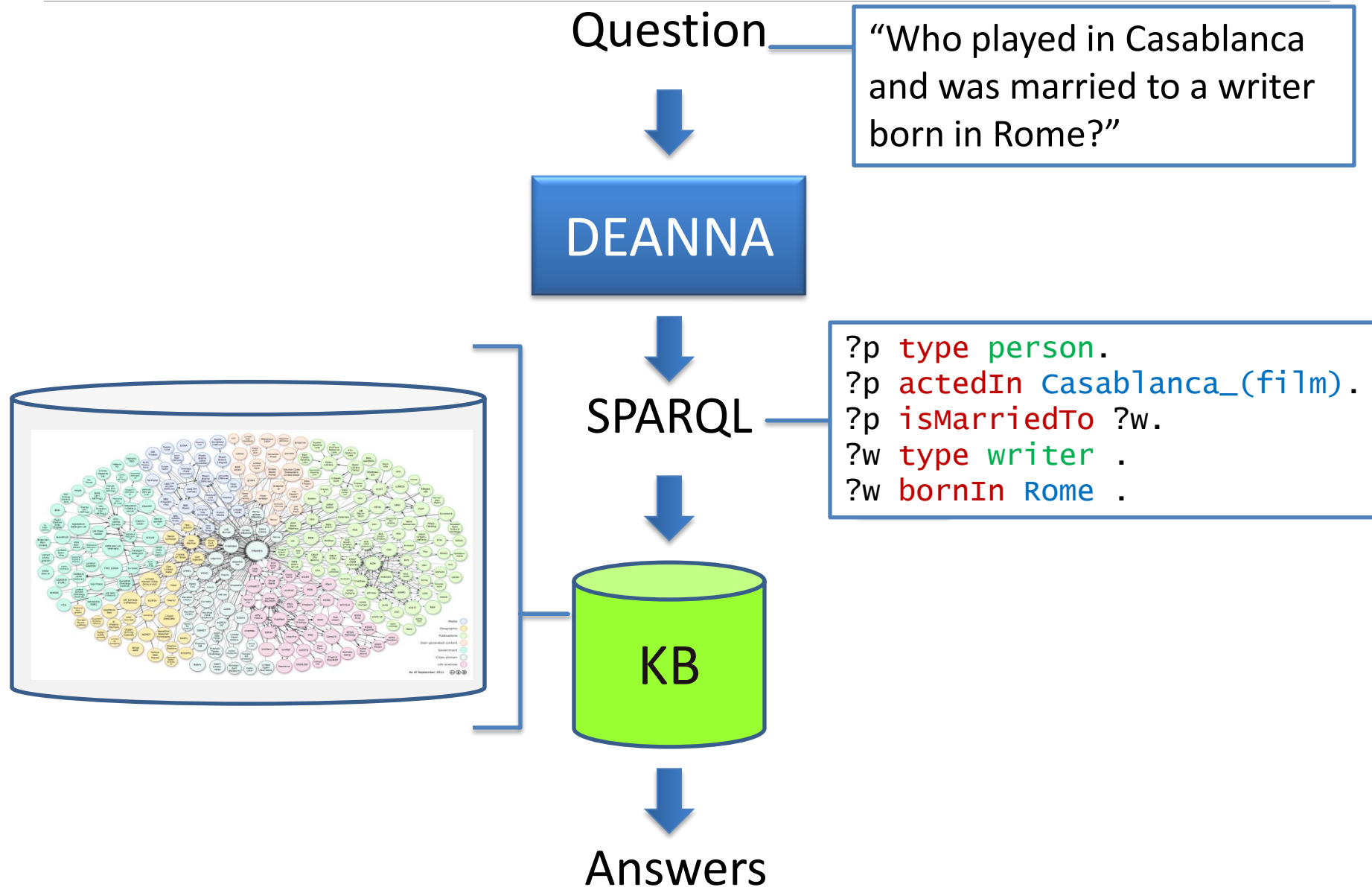
What is DEANNA?



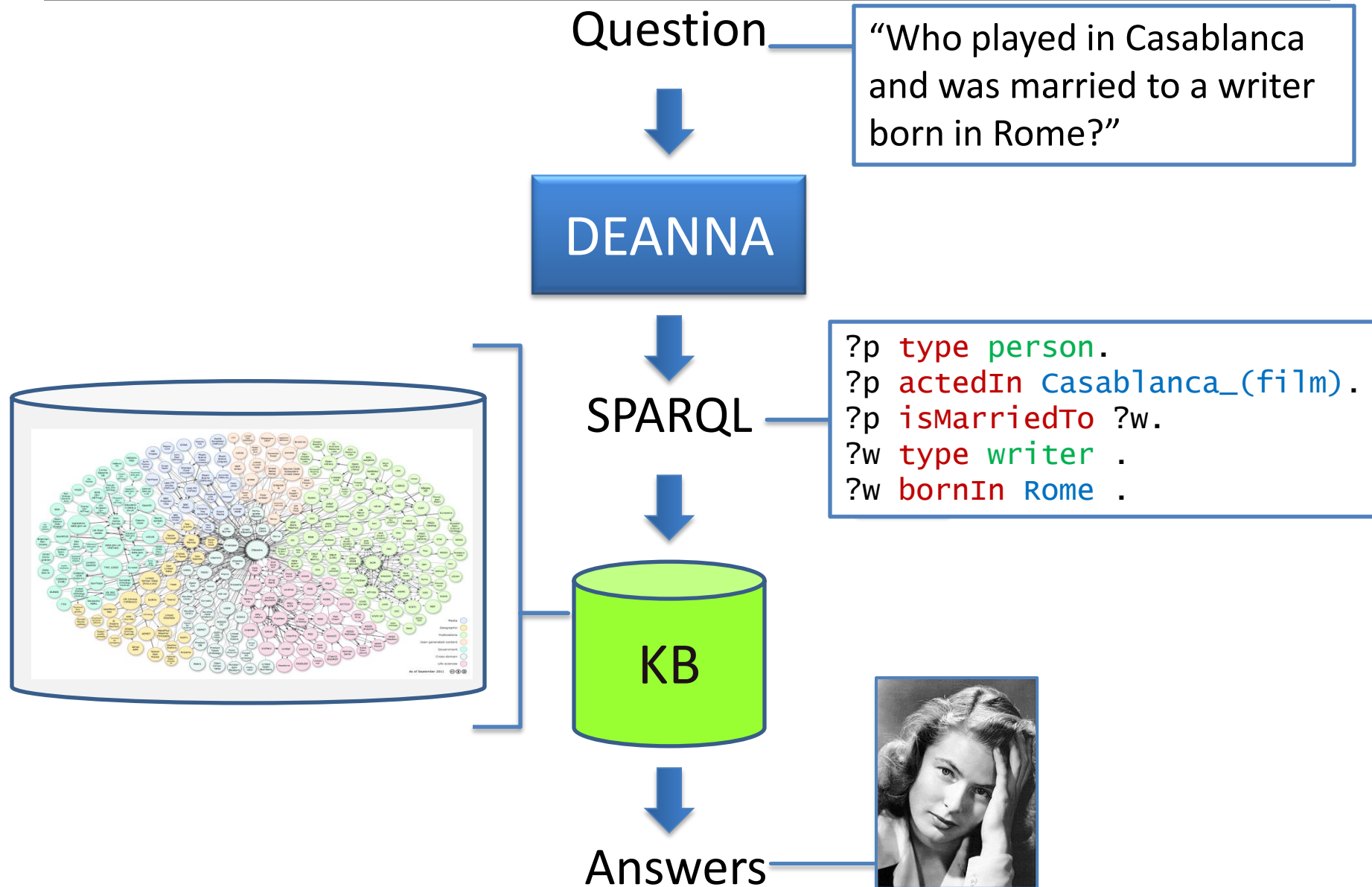
What is DEANNA?



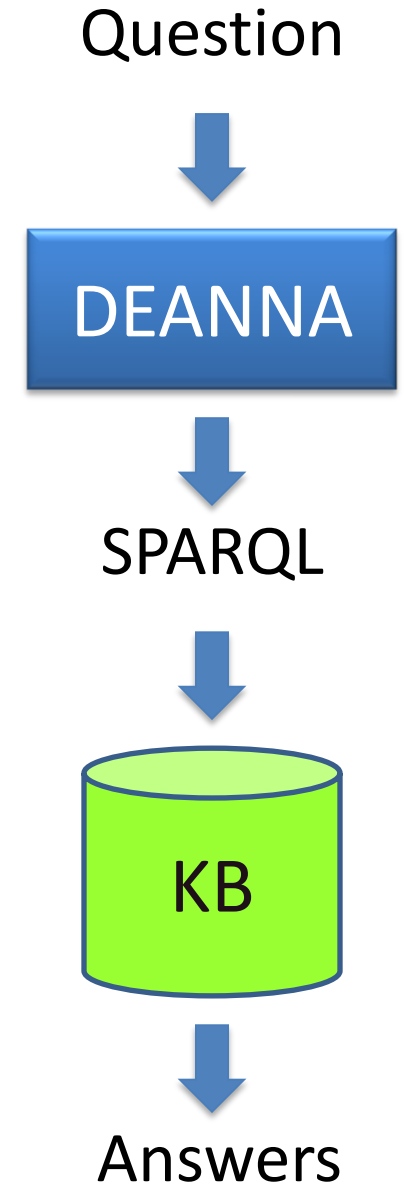
What is DEANNA?



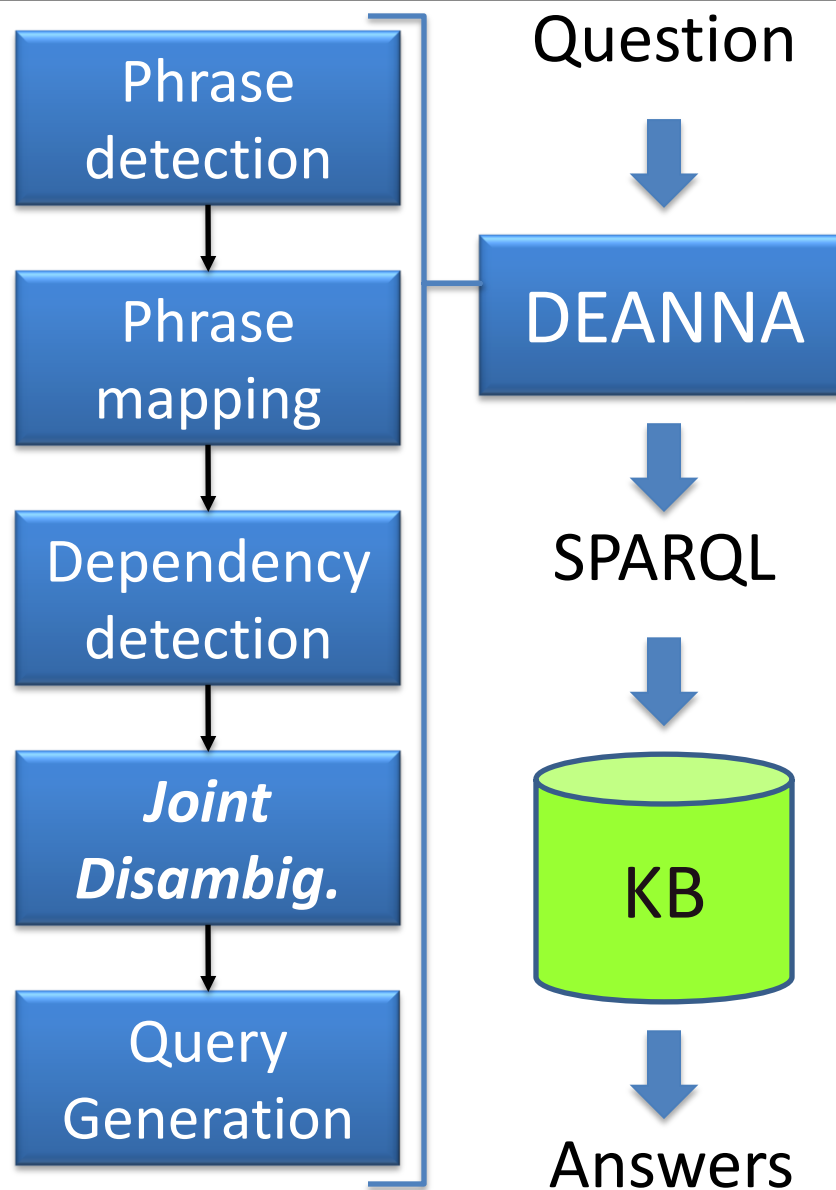
What is DEANNA?



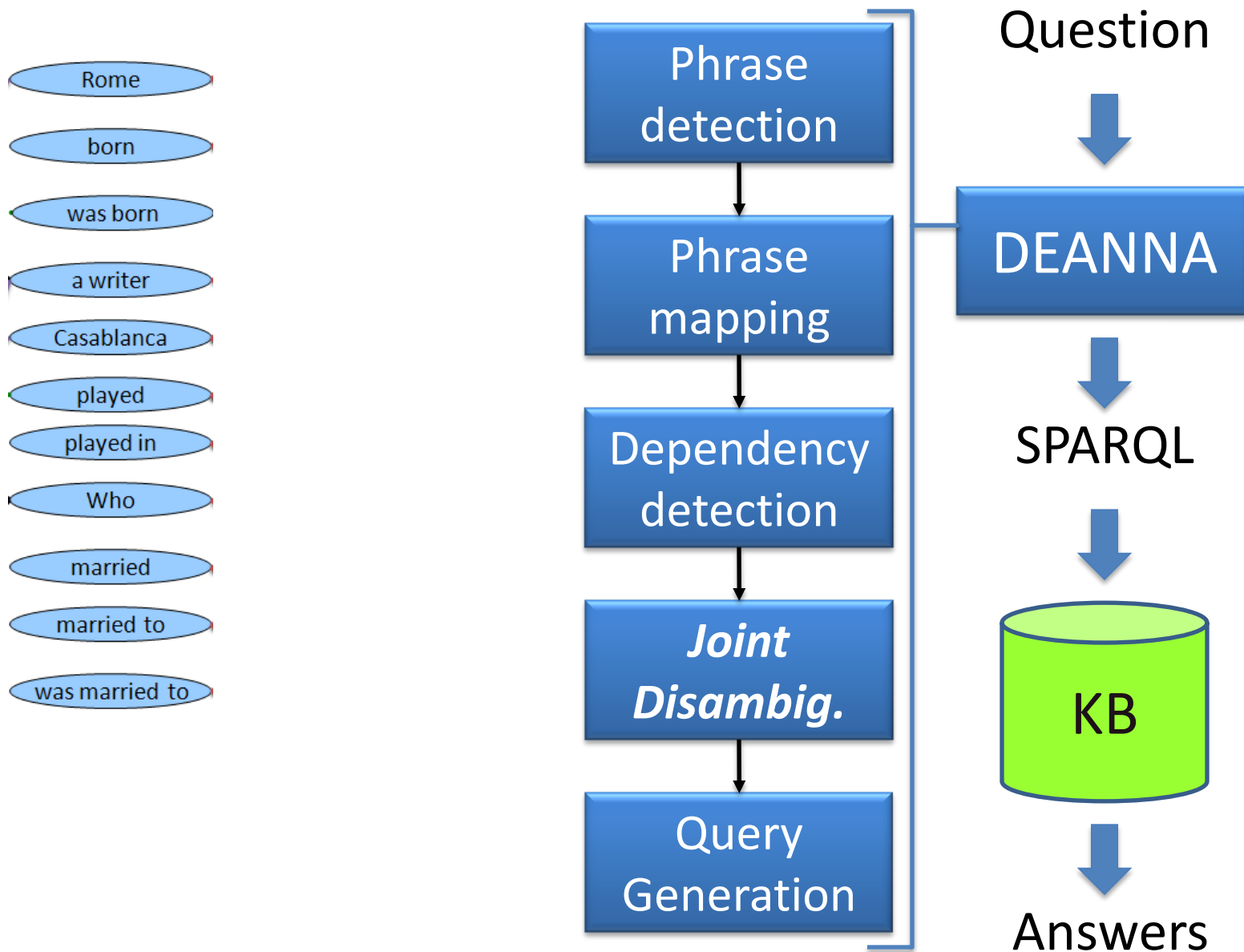
Inside DEANNA



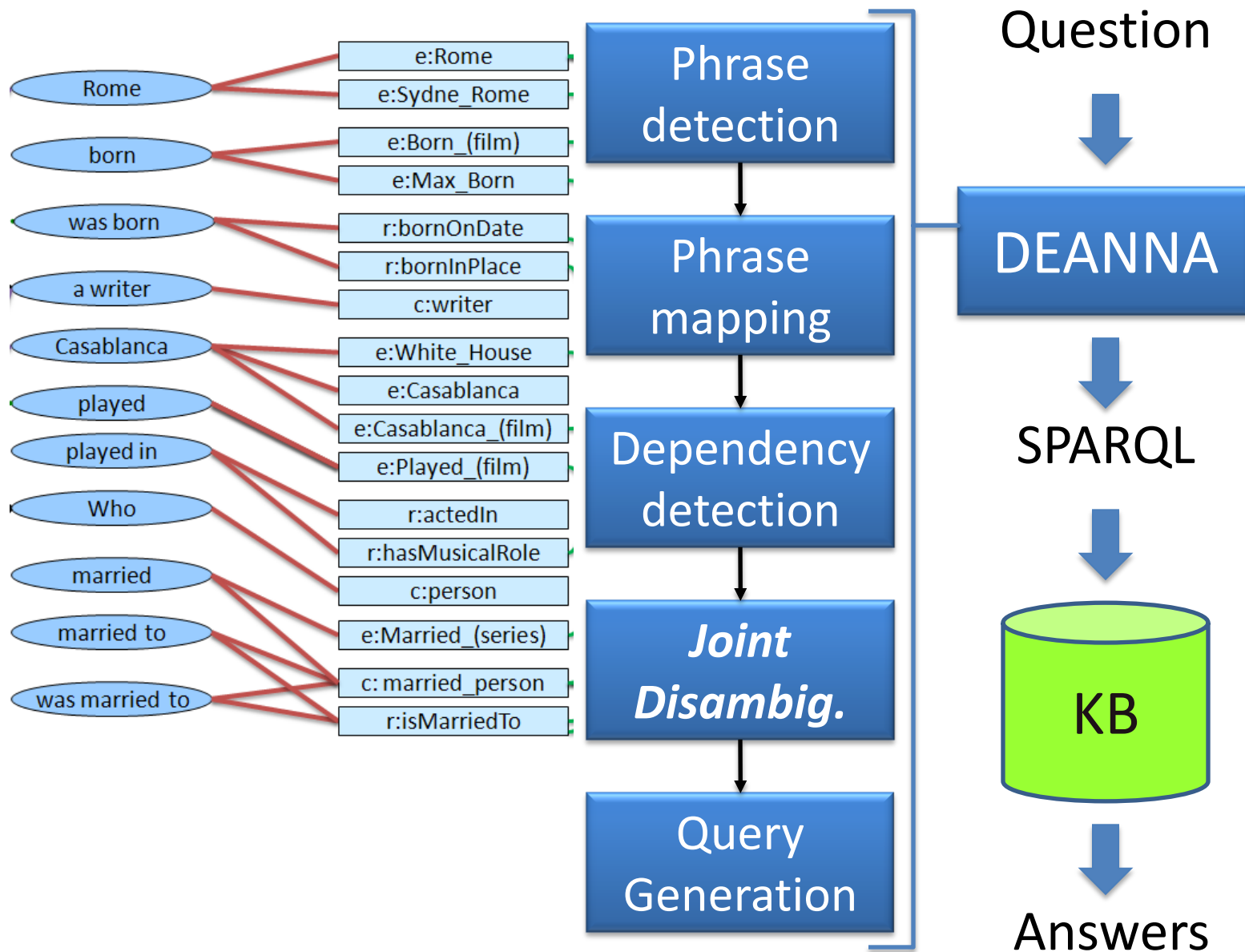
Inside DEANNA



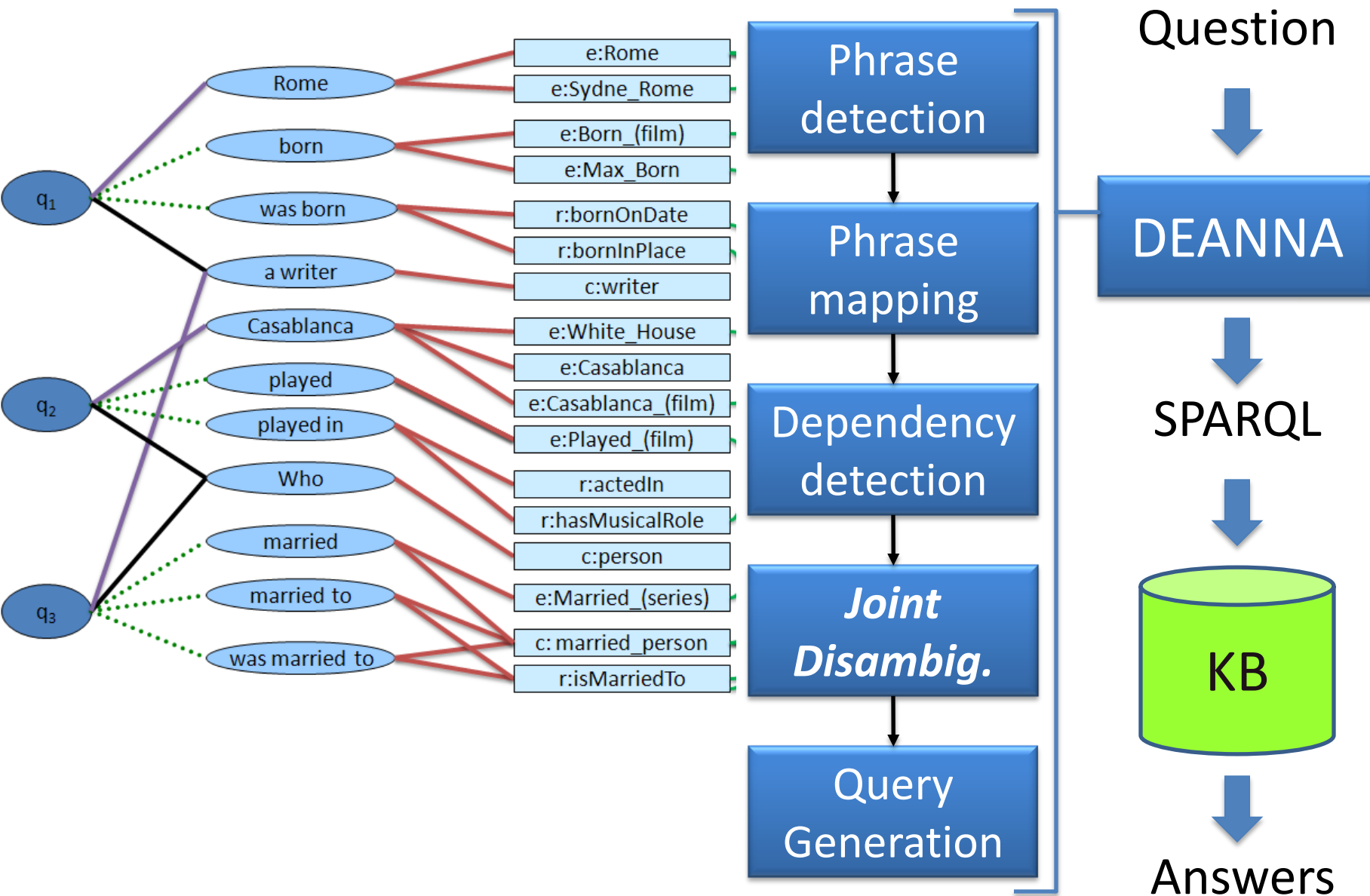
Inside DEANNA



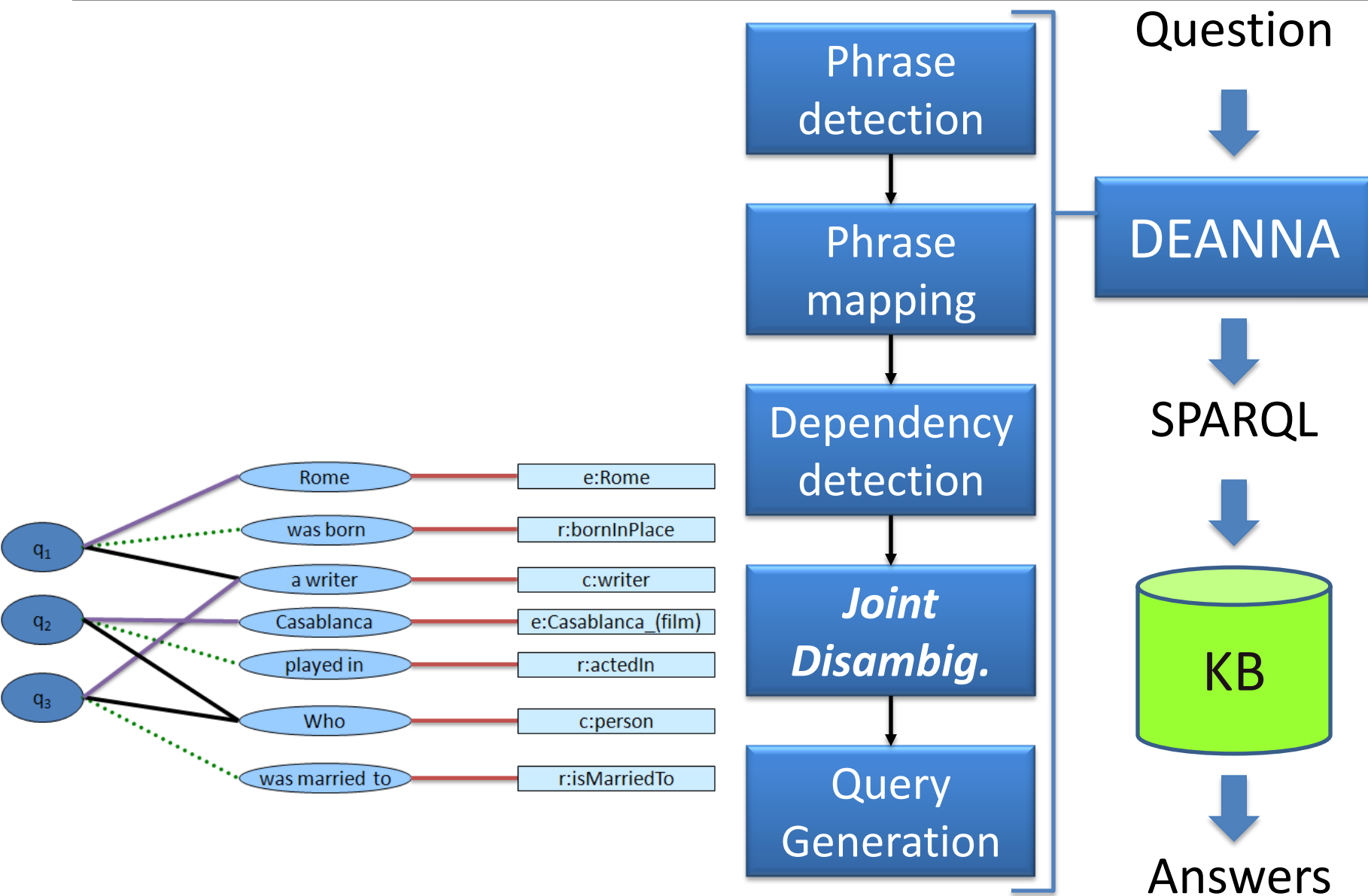
Inside DEANNA



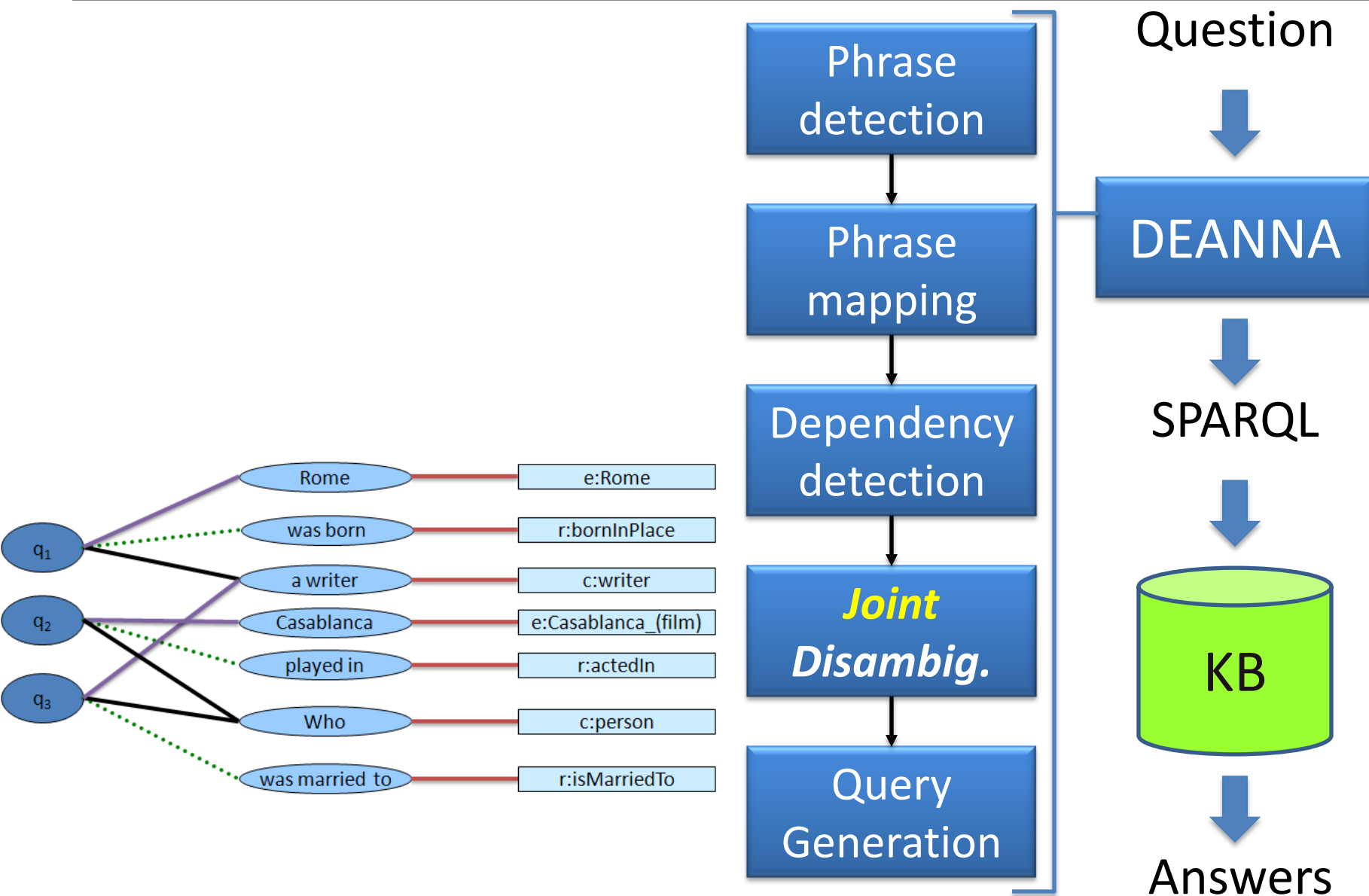
Inside DEANNA



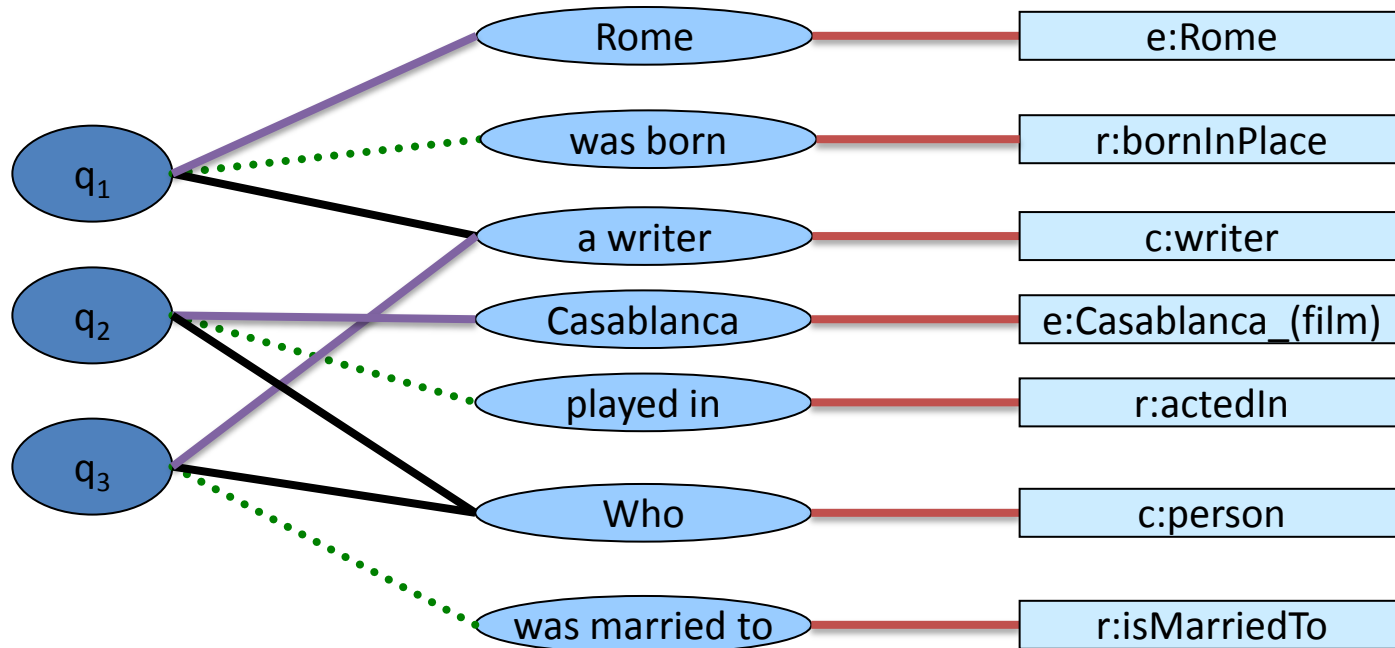
Inside DEANNA



Inside DEANNA

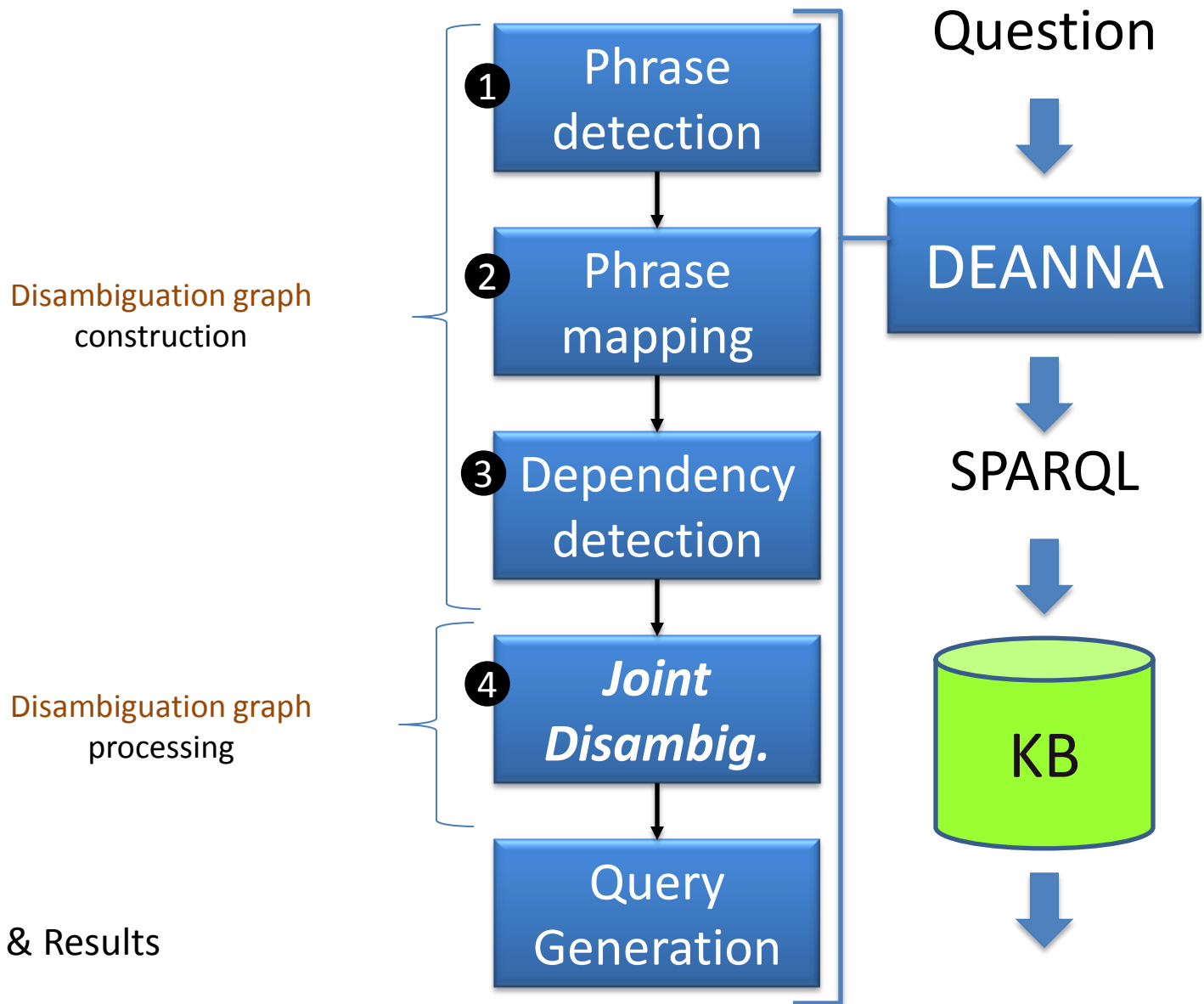


Structured Query Generation



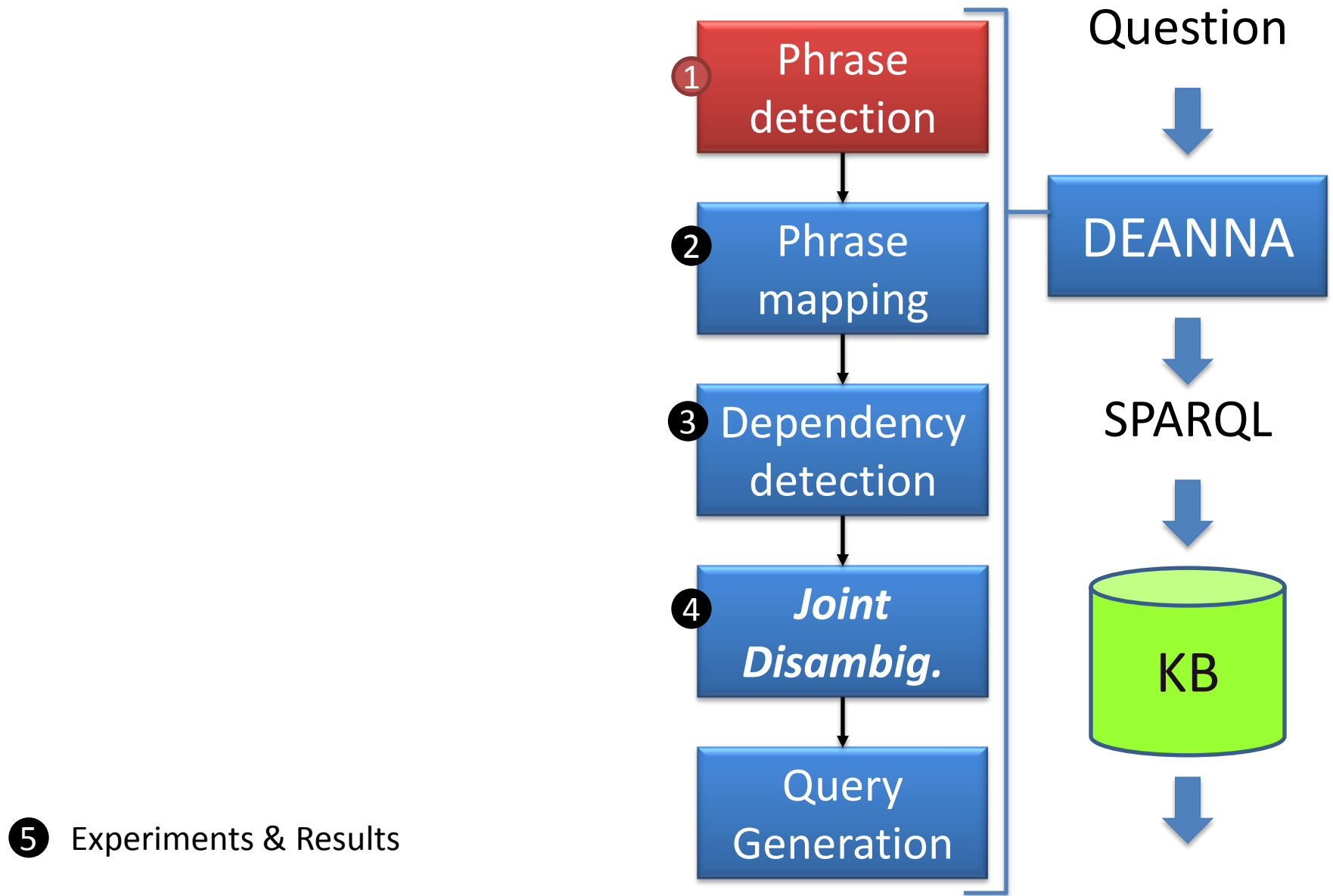
```
SELECT ?p WHERE {  
  ?p type person .  
  ?p actedIn Casablanca_(film) .  
  ?p isMarriedTo ?w .  
  ?w type writer .  
  ?w bornIn Rome }
```

Outline



5 Experiments & Results

Outline



5 Experiments & Results

Phrase Detection

- **Concepts:** entities & classes:

Dictionary-based

Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
...	...

- **Relations:**

Mainly use **Reverb** [Fader et al.

EMNLP'11]: $V \mid VP \mid VW^*P$

... was/VBD married/VBN to/TO a/DT...

Phrase Detection

a writer

Casablanca

played

played in

Who

married

married to

was married to

- **Concepts:** entities & classes:

Dictionary-based

Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
...	...

- **Relations:**

Mainly use **Reverb** [Fader et al.

EMNLP'11]: $V \mid VP \mid VW^*P$

... was/VBD married/VBN to/TO a/DT...

Phrase Detection

a writer

Casablanca

played

played in

Who

married

married to

was married to

- **Concepts:** entities & classes:

Dictionary-based

Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
...	...

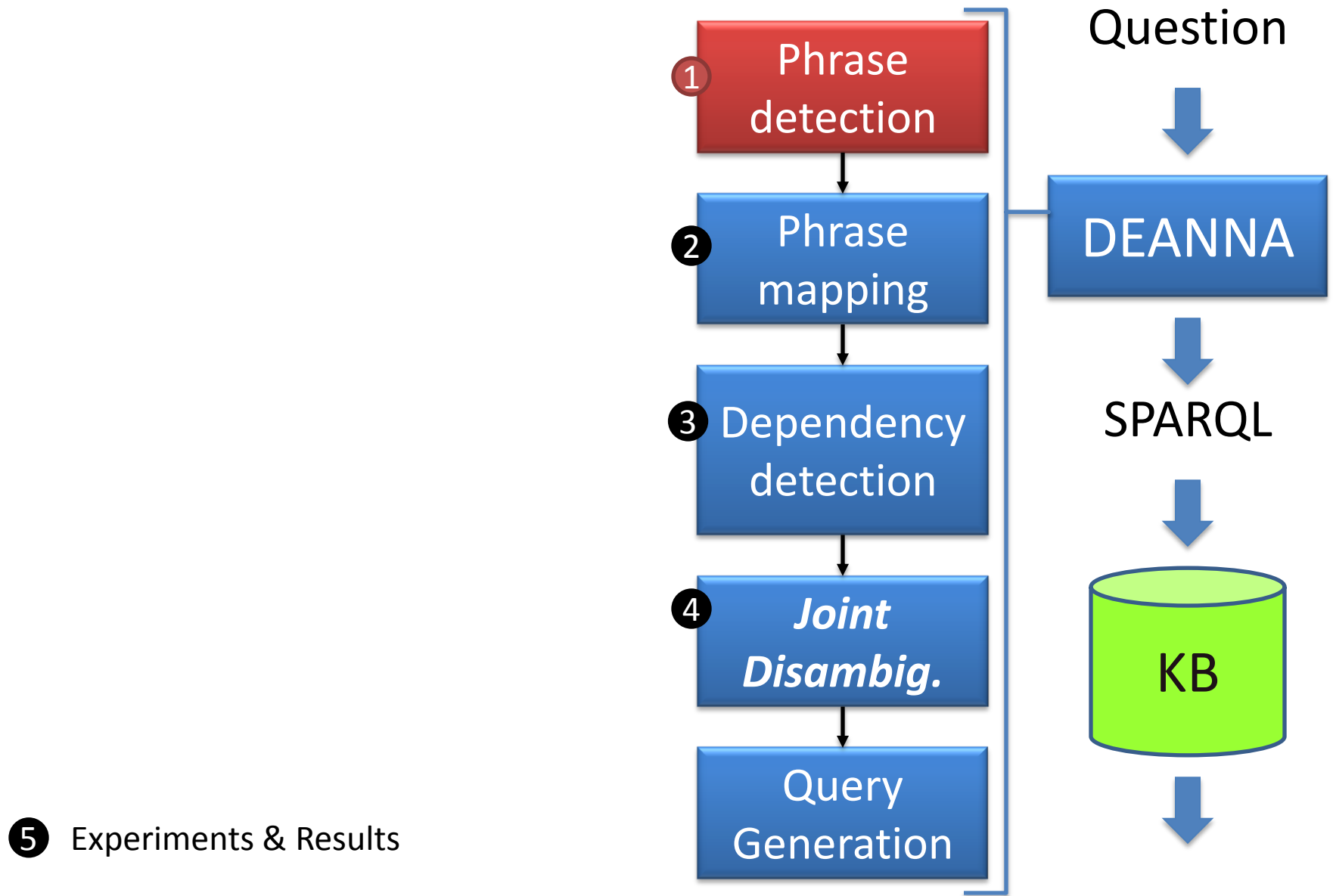
- **Relations:**

Mainly use **Reverb** [Fader et al.

EMNLP'11]: $V \mid VP \mid VW^*P$

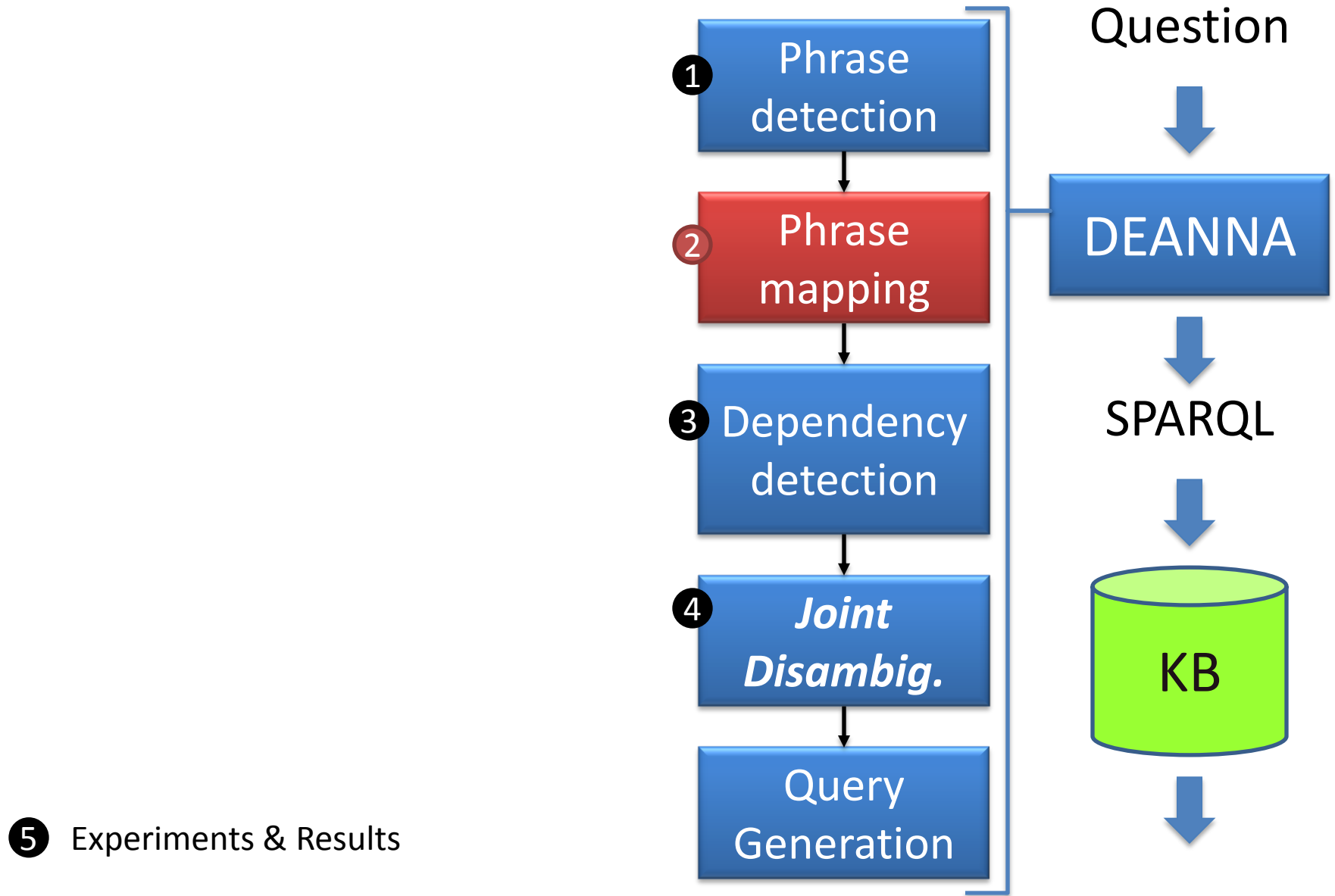
... ***was/VBD married/VBN to/TO a/DT...***

Outline



5 Experiments & Results

Outline



5 Experiments & Results

Phrase Mapping

Casablanca

played

played in

Phrase Mapping

Concepts: entities & classes:
Dictionary-based

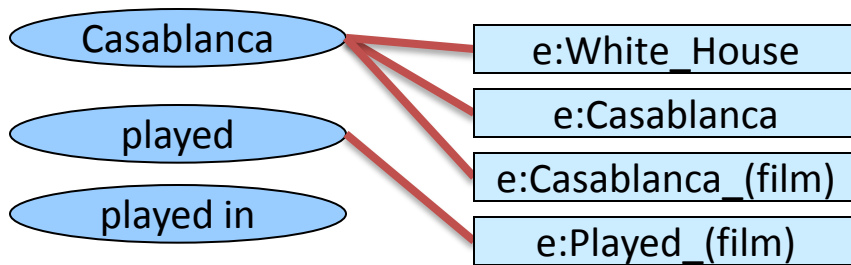
Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
Played_(film)	Played

Casablanca

played

played in

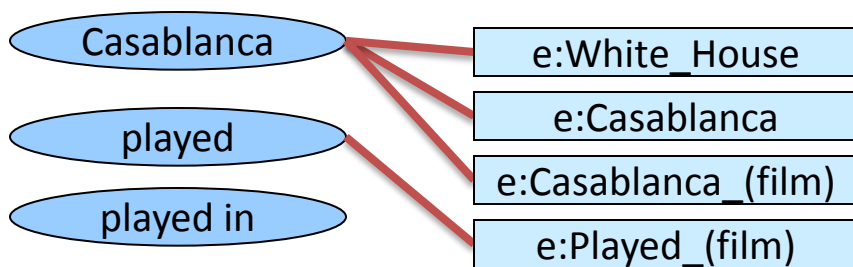
Phrase Mapping



Concepts: entities & classes:
Dictionary-based

Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
Played_(film)	Played

Phrase Mapping



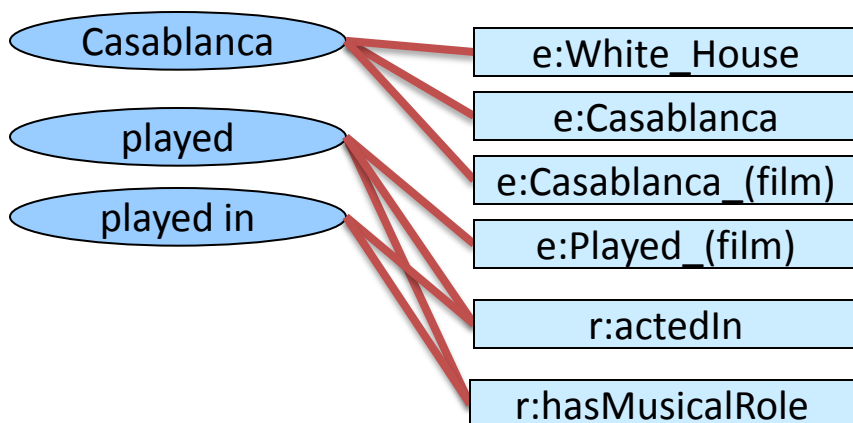
Concepts: entities & classes:
Dictionary-based

Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
Played_(film)	Played

•

Relation	Phrase
actedIn	acted in
actedIn	played in
hasMusicalRole	plays
hasMusicalRole	mastered

Phrase Mapping



Concepts: entities & classes:
Dictionary-based

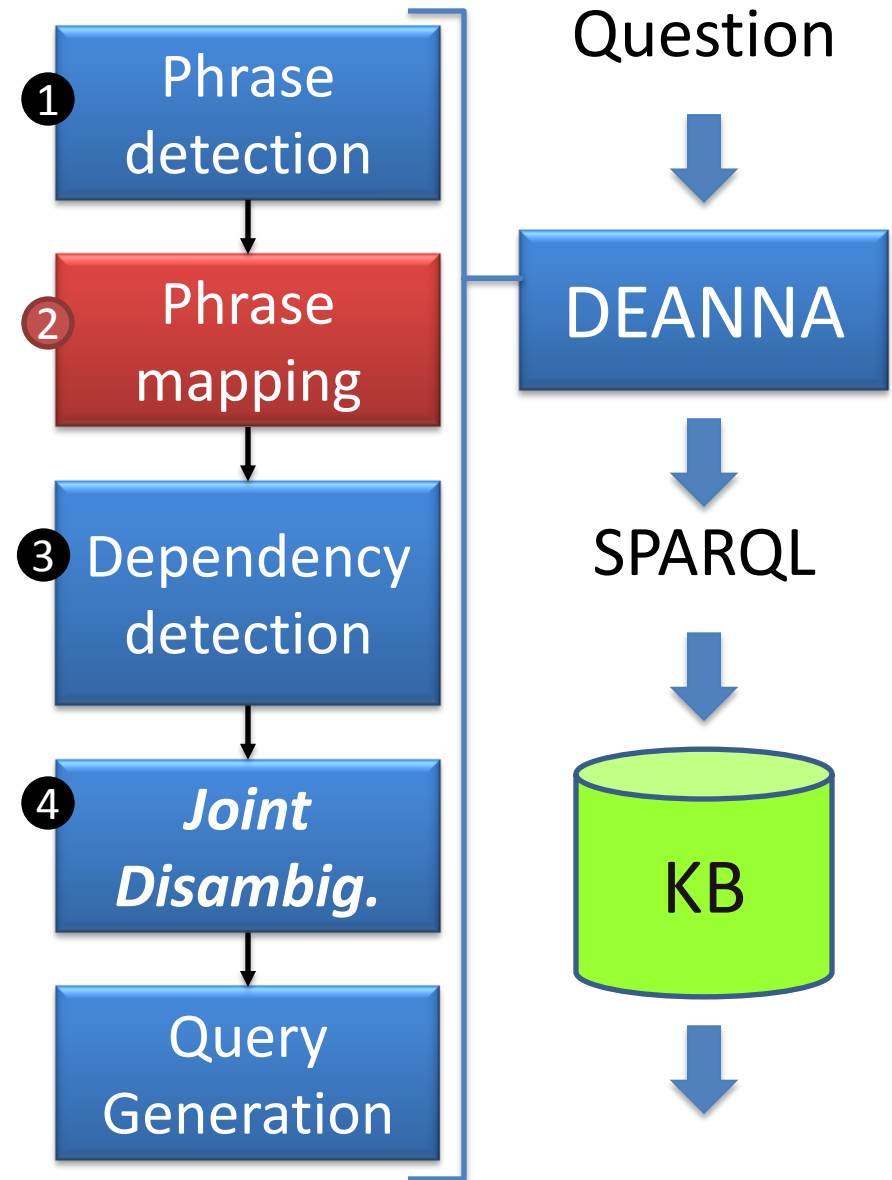
Concept	Phrase
Casablanca	Casablanca
Casablanca	Casablanca, Morocco
Casablanca_(film)	Casablanca the film
Casablanca_(film)	Casablanca
Played_(film)	Played

•

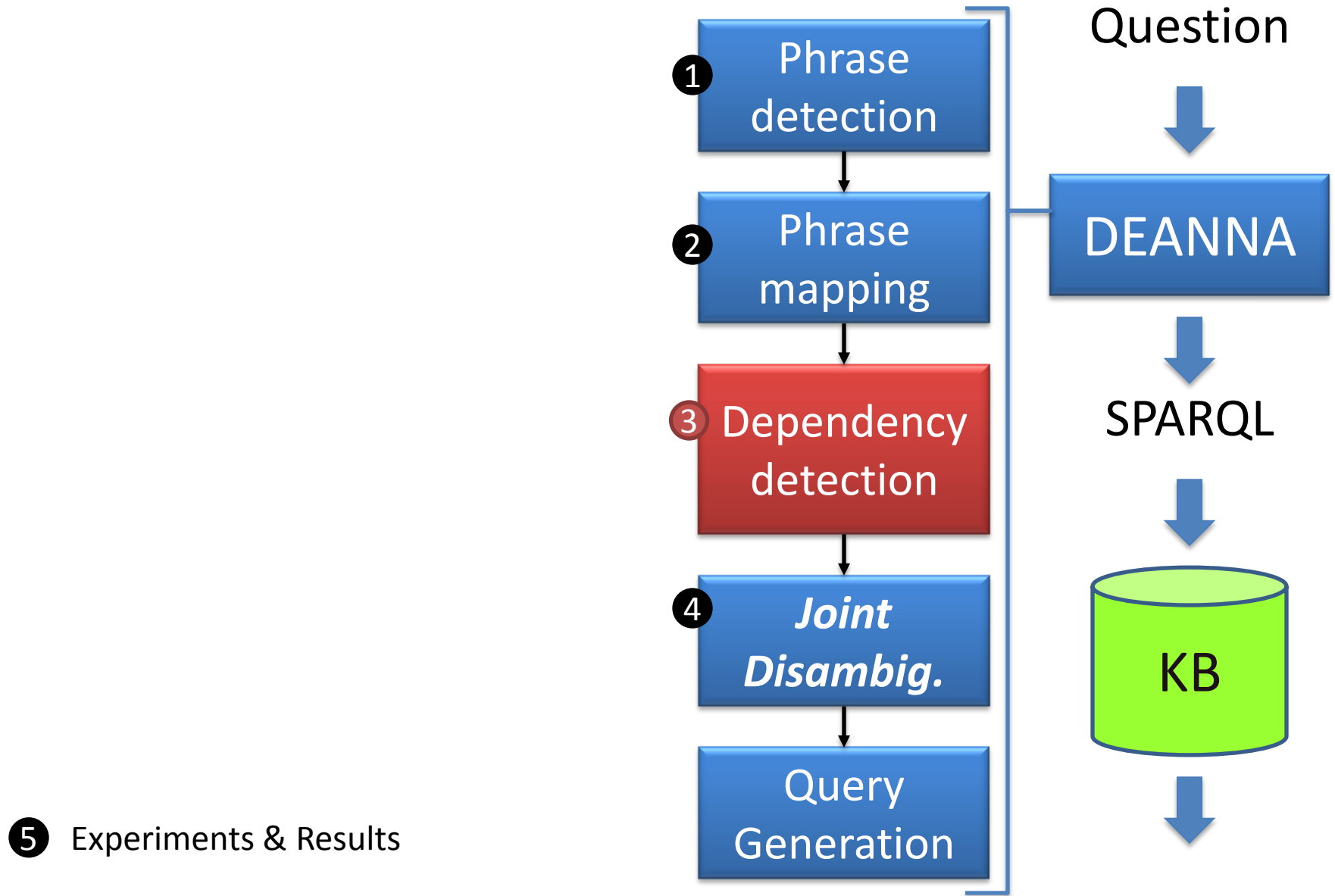
Relation	Phrase
actedIn	acted in
actedIn	played in
hasMusicalRole	plays
hasMusicalRole	mastered

Outline

5 Experiments & Results

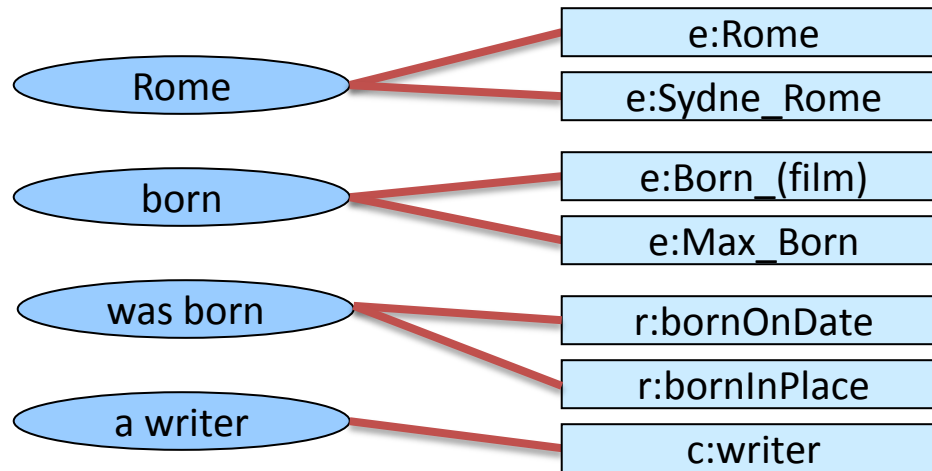


Outline

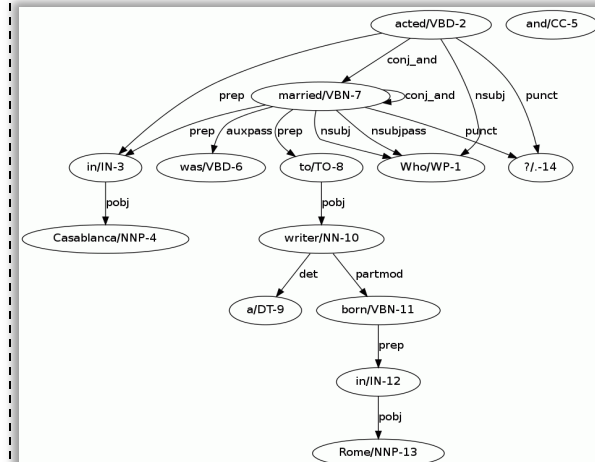


5 Experiments & Results

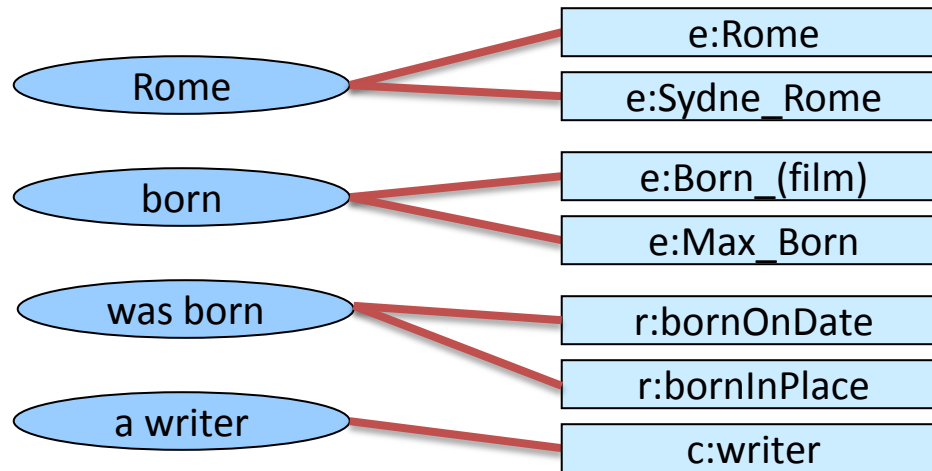
Dependency Detection



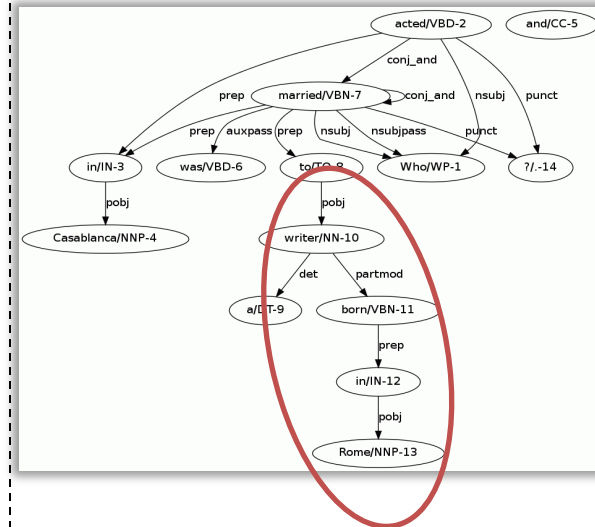
Look for specific patterns in
dependency parses
[de Marneffe et al. LREC'06]



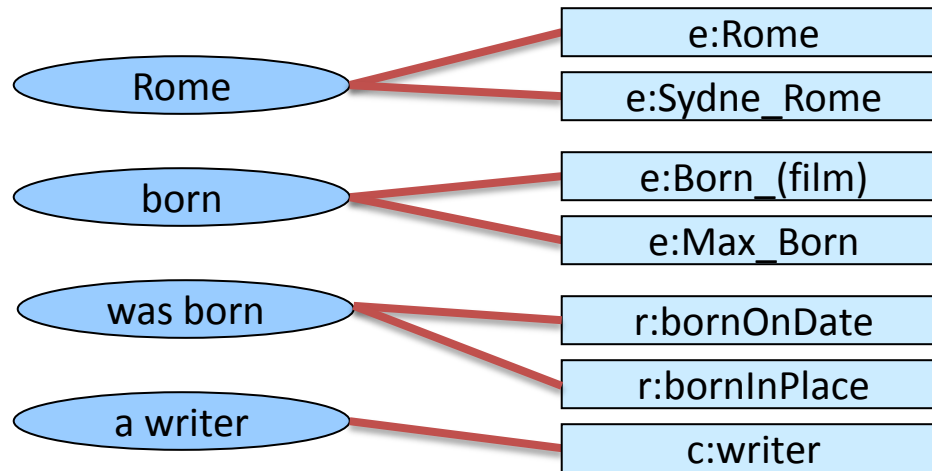
Dependency Detection



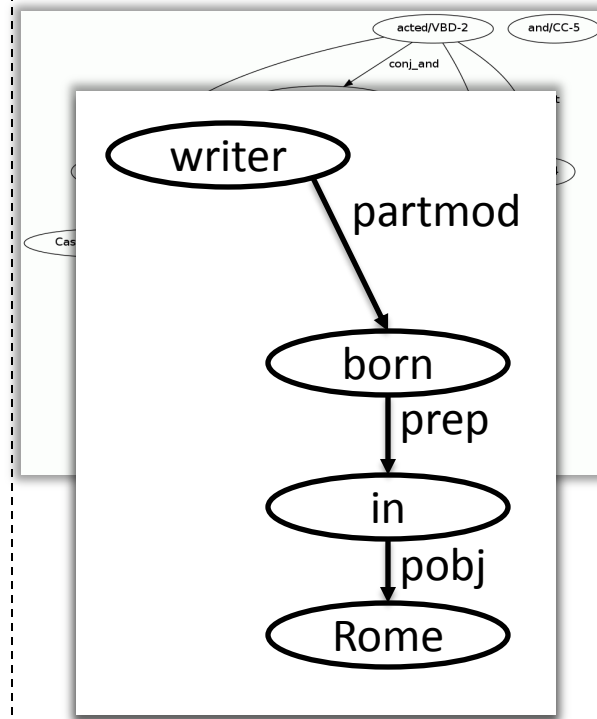
Look for specific patterns in
dependency parses
[de Marneffe et al. LREC'06]



Dependency Detection

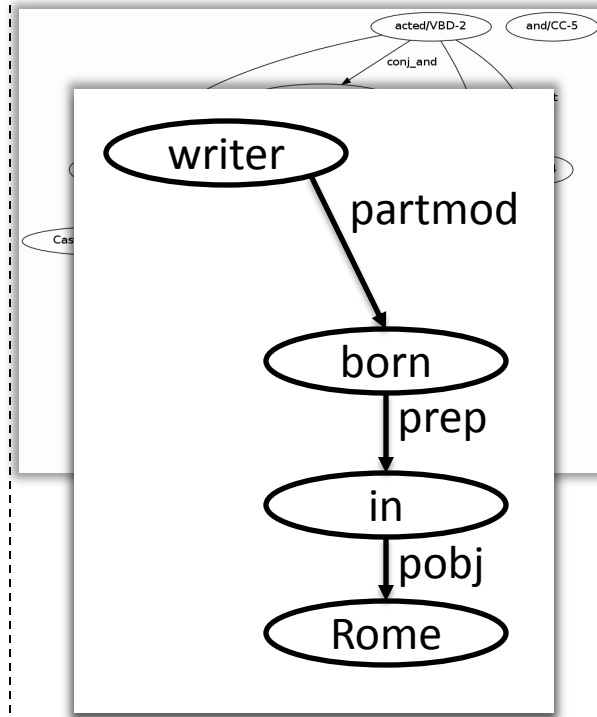
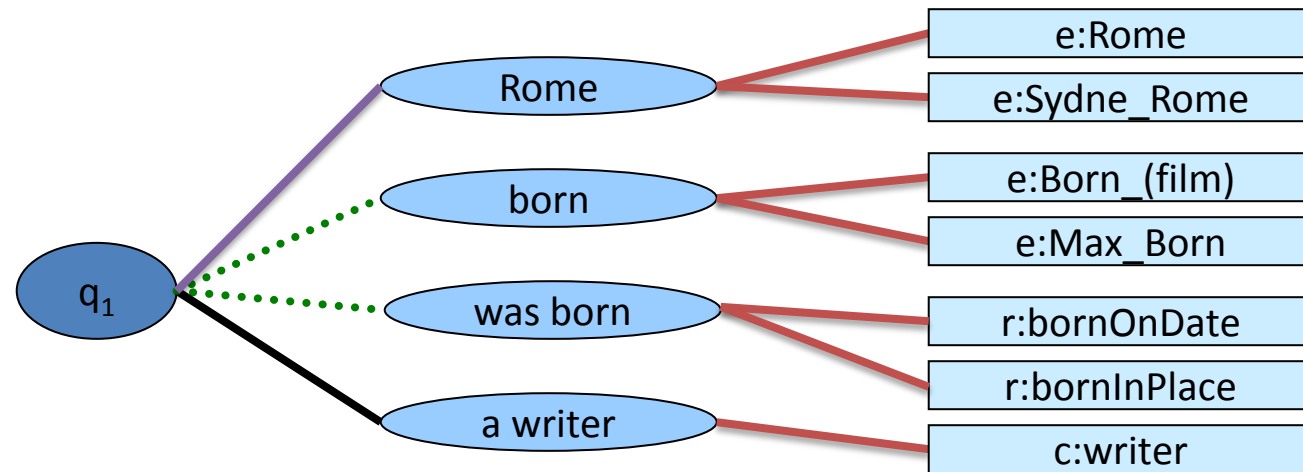


Look for specific patterns in
dependency parses
[de Marneffe et al. LREC'06]



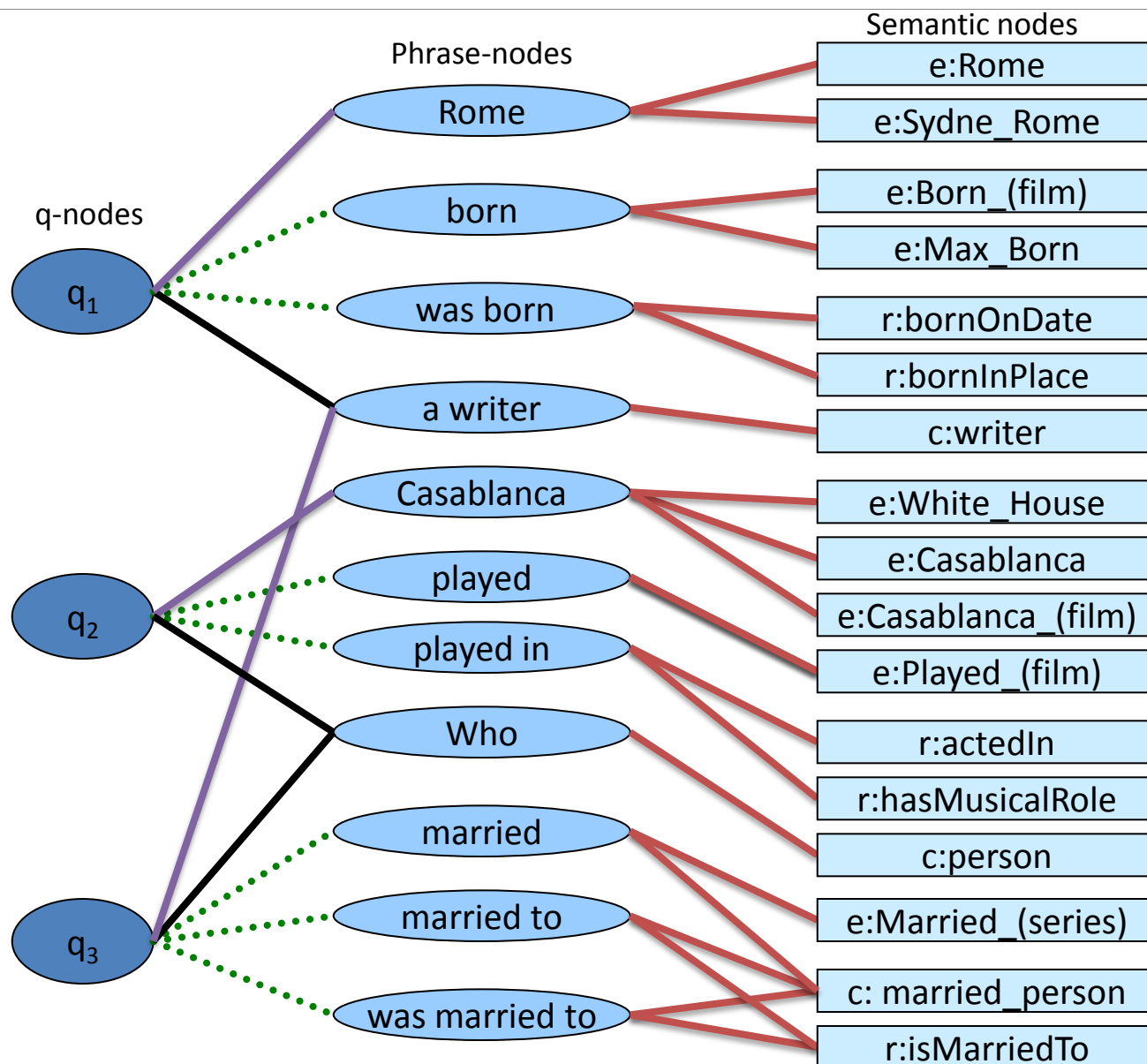
Dependency Detection

Look for specific patterns in
dependency parses
[de Marneffe et al.
LREC'06]



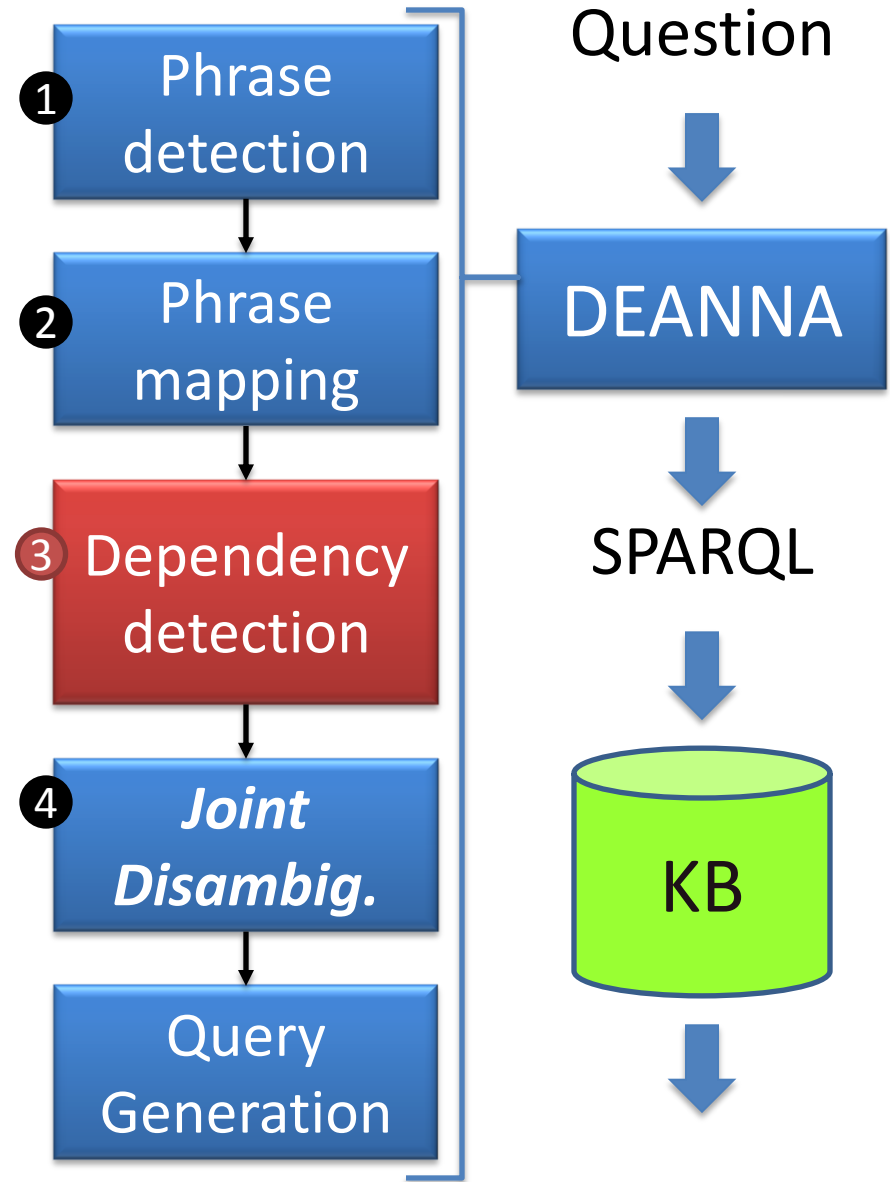
RESULT

Disambiguation Graph

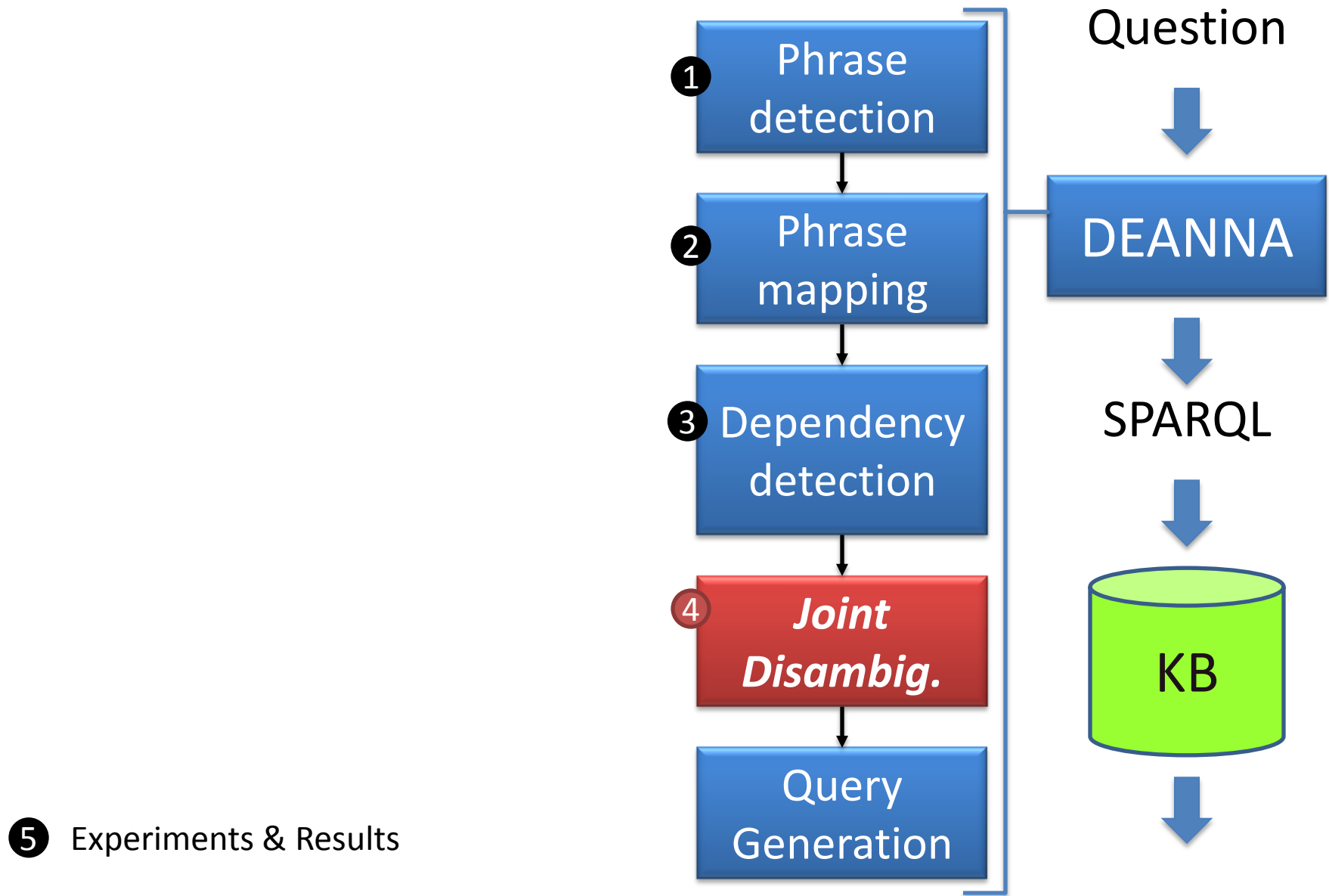


Outline

5 Experiments & Results

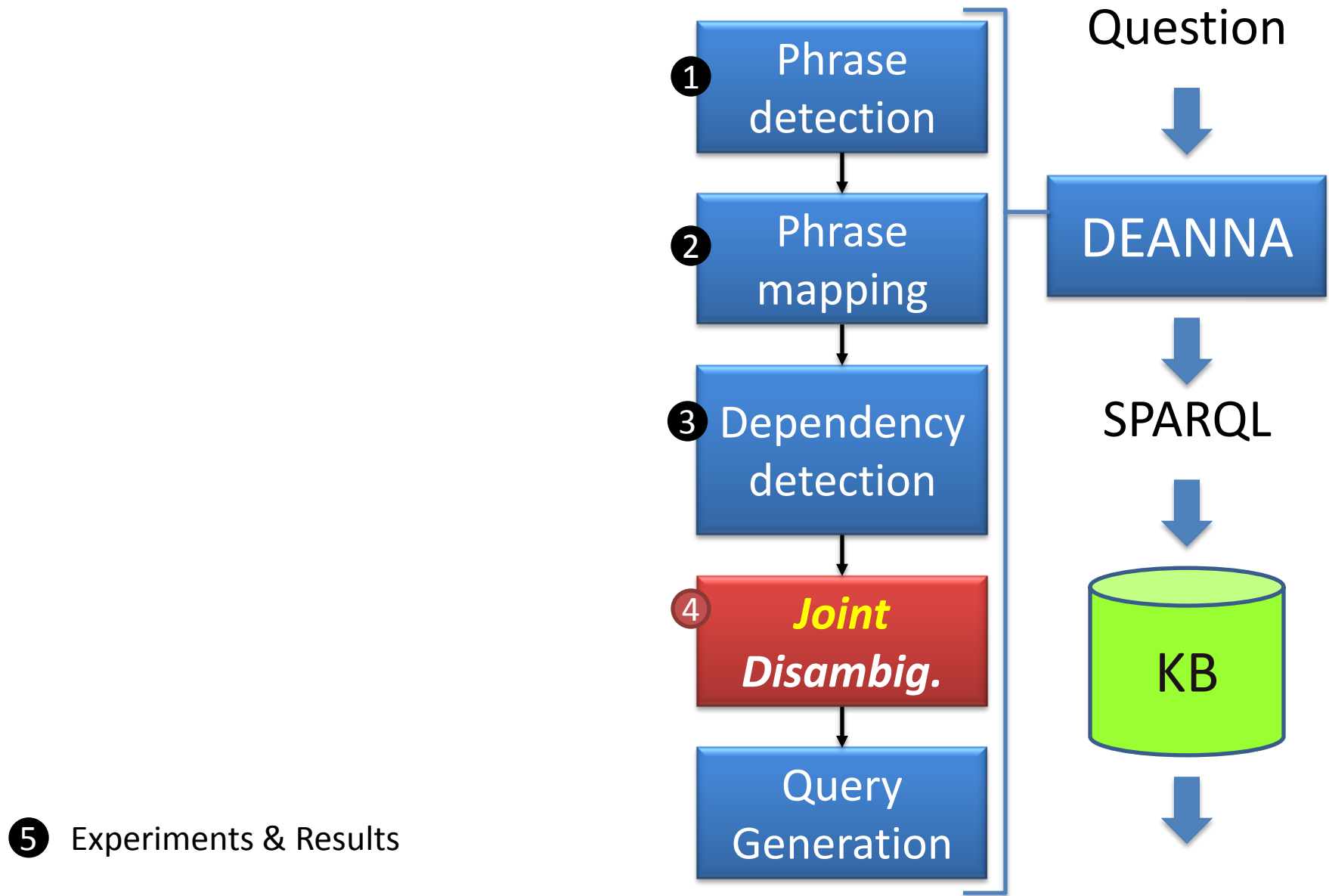


Outline



5 Experiments & Results

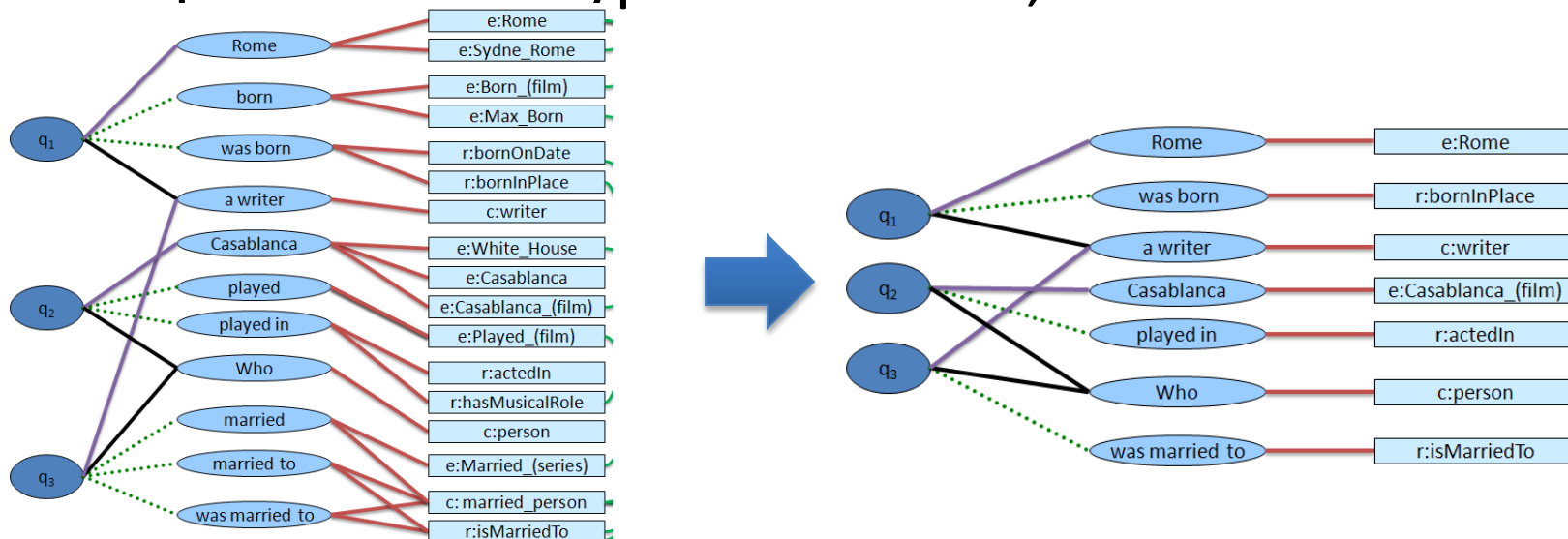
Outline



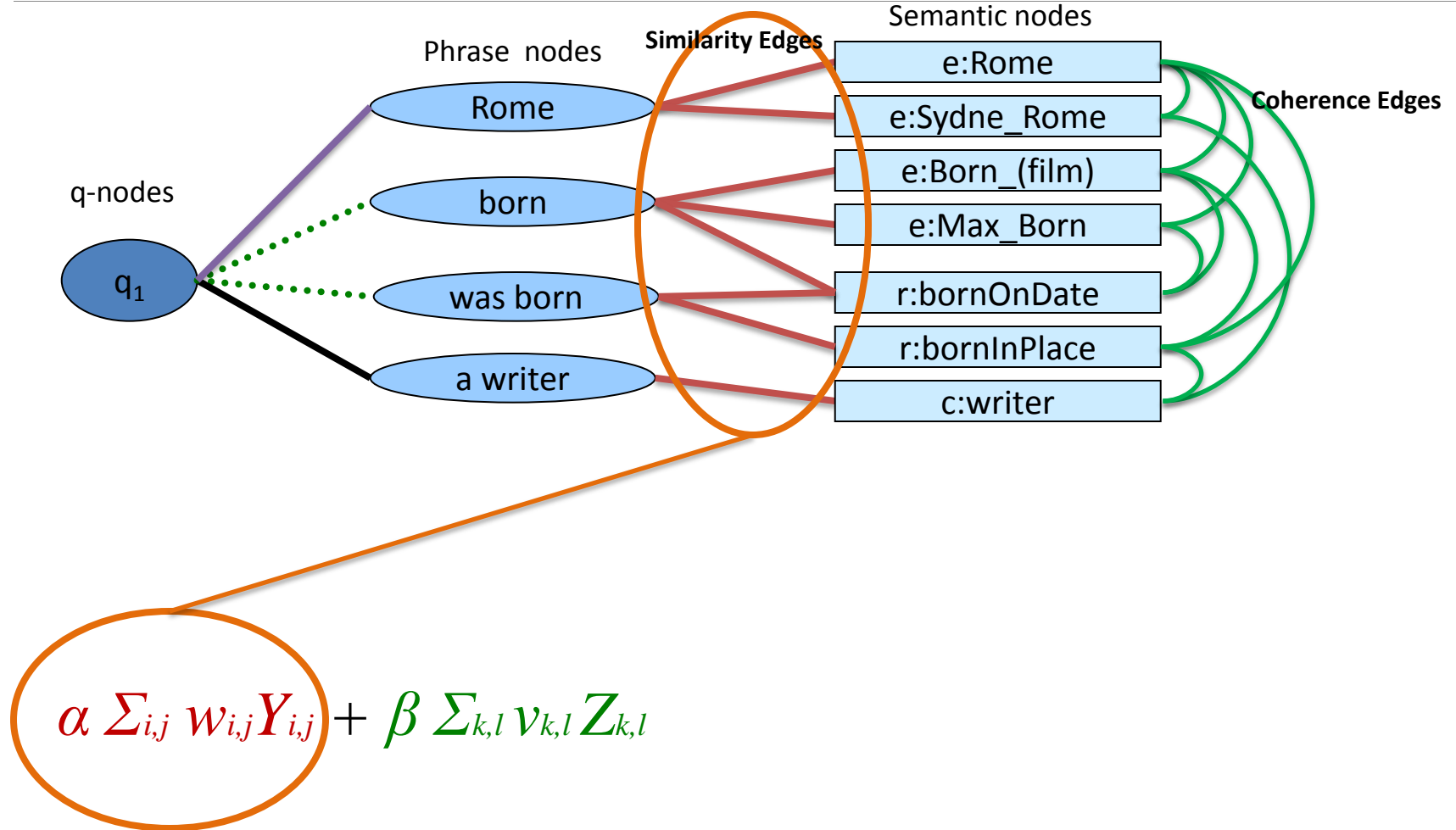
5 Experiments & Results

Joint Disambiguation - ILP

- *ILP*: Integer Linear Programming
- *maximize* $\alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} v_{k,l} Z_{k,l} + \dots$
- Subject to:
 - No token in multiple phrases,
 - Triples observe type constraints, ...

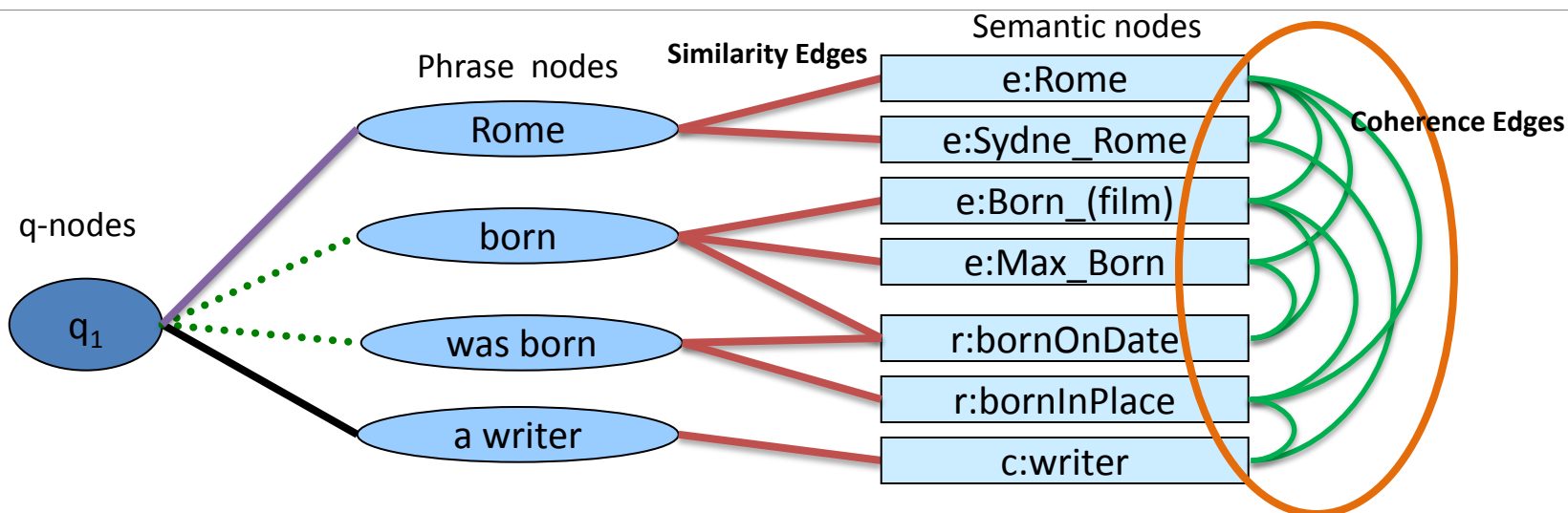


Joint Disambiguation – Objective



Prior

Joint Disambiguation – Objective

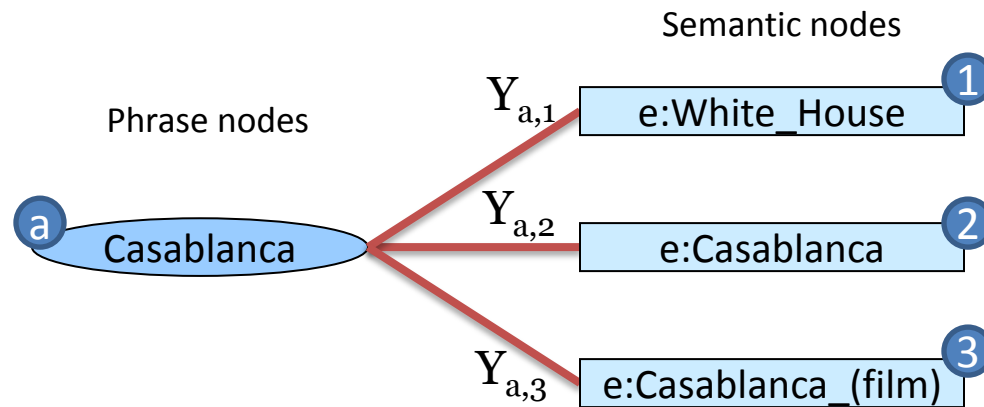


$$\alpha \sum_{i,j} w_{i,j} Y_{i,j} + \beta \sum_{k,l} v_{k,l} Z_{k,l}$$

Coherence

Joint Disambiguation – Constraints

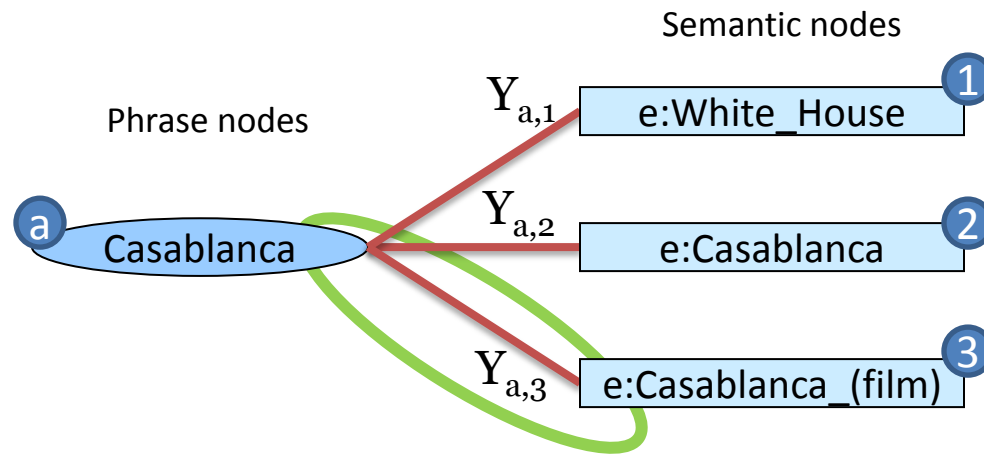
A phrase node can be assigned to one semantic node:



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

Joint Disambiguation – Constraints

A phrase node can be assigned to one semantic node:



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

Joint Disambiguation – Constraints

Classes translate to **type-constrained variables**

→ Every semantic triple should have a class to
join & project!

person actedIn Casablanca_(film)



?x type person . ?x actedIn Casablanca_(film)

Joint Disambiguation – Constraints

Classes translate to **type-constrained variables**

→ Every semantic triple should have a class to
join & project!

person actedIn Casablanca_(film)

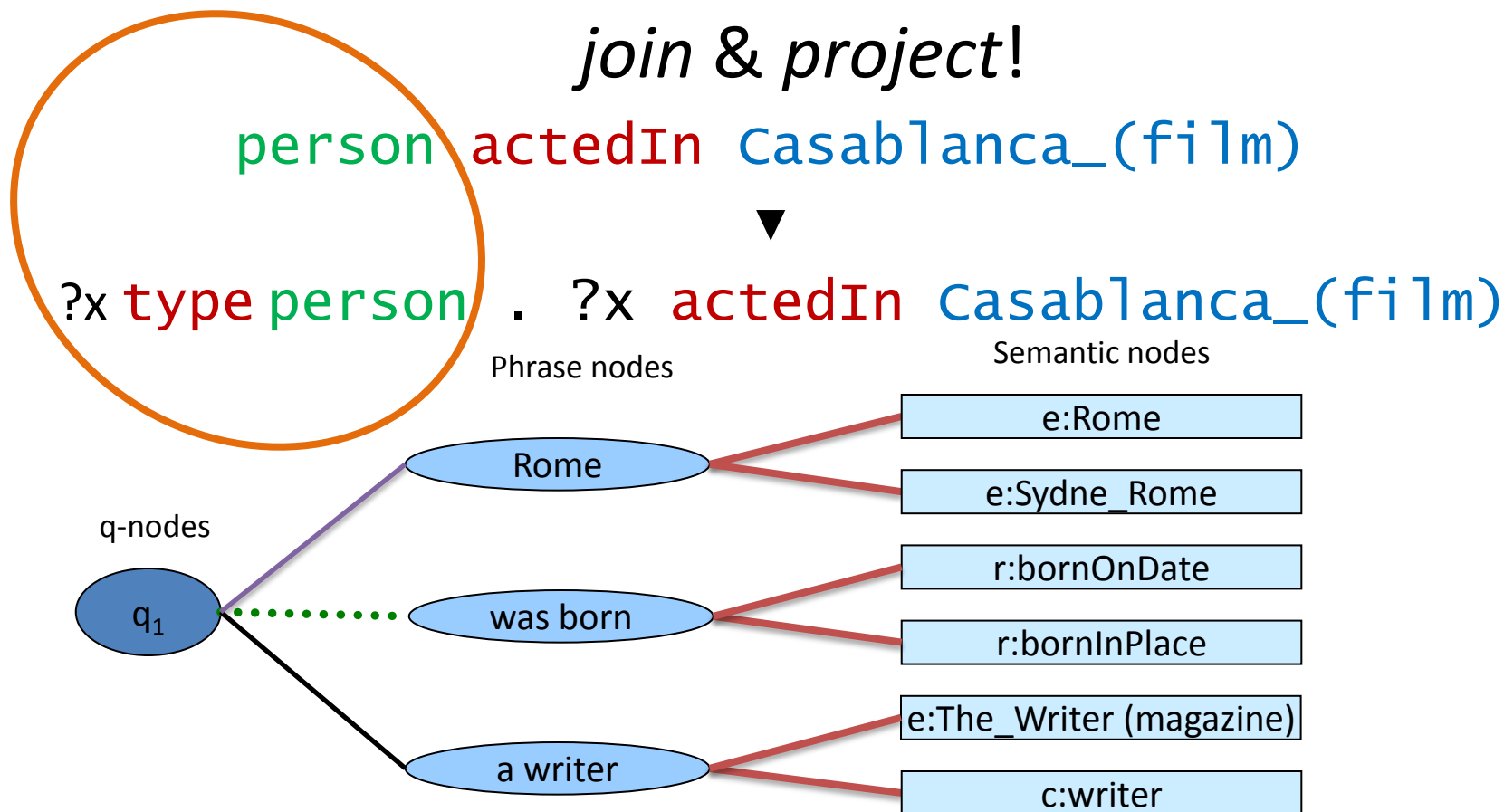


?x type person . ?x actedIn Casablanca_(film)

Joint Disambiguation – Constraints

Classes translate to **type-constrained variables**

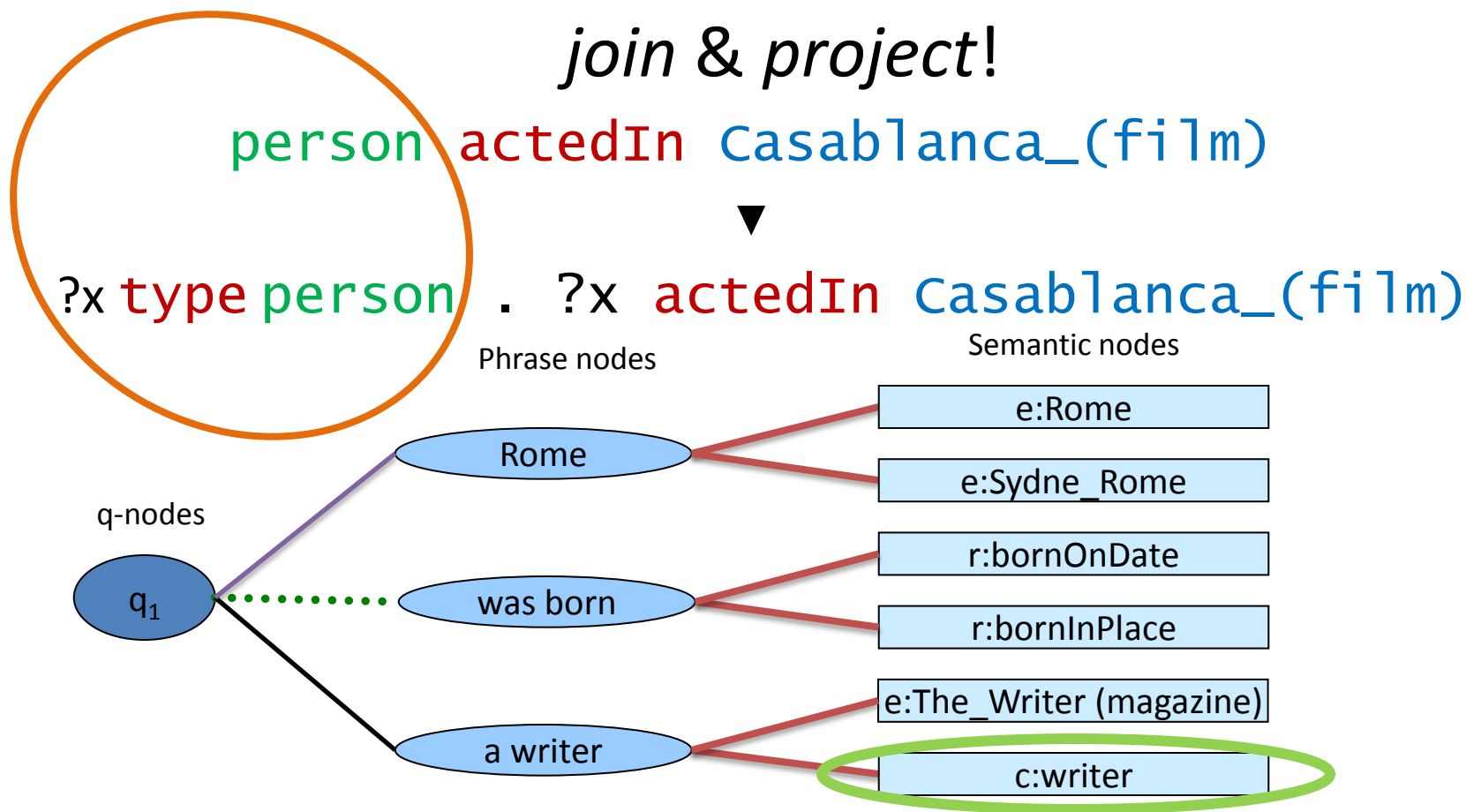
→ Every semantic triple should have a class to *join & project*!



Joint Disambiguation – Constraints

Classes translate to **type-constrained variables**

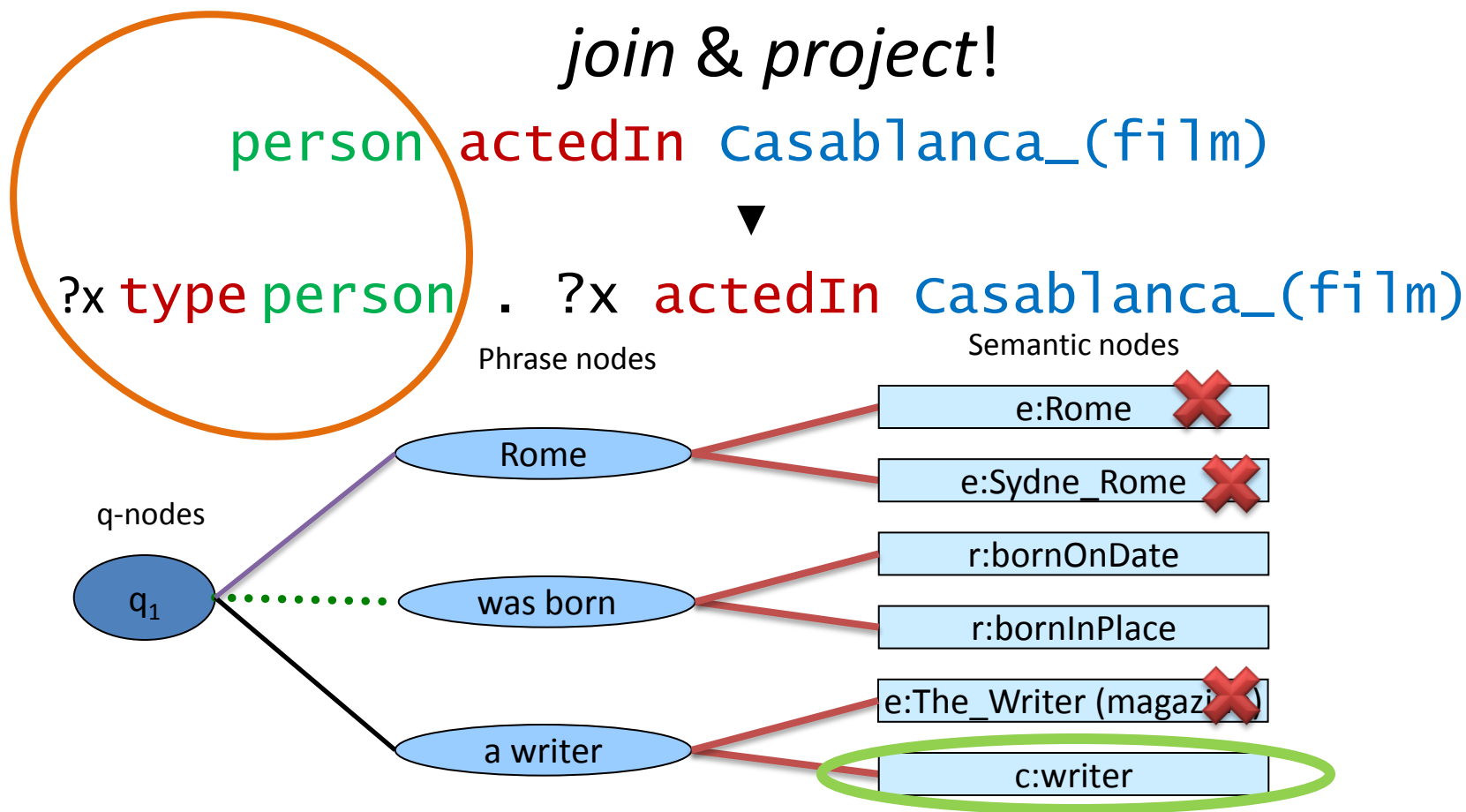
→ Every semantic triple should have a class to *join & project*!



Joint Disambiguation – Constraints

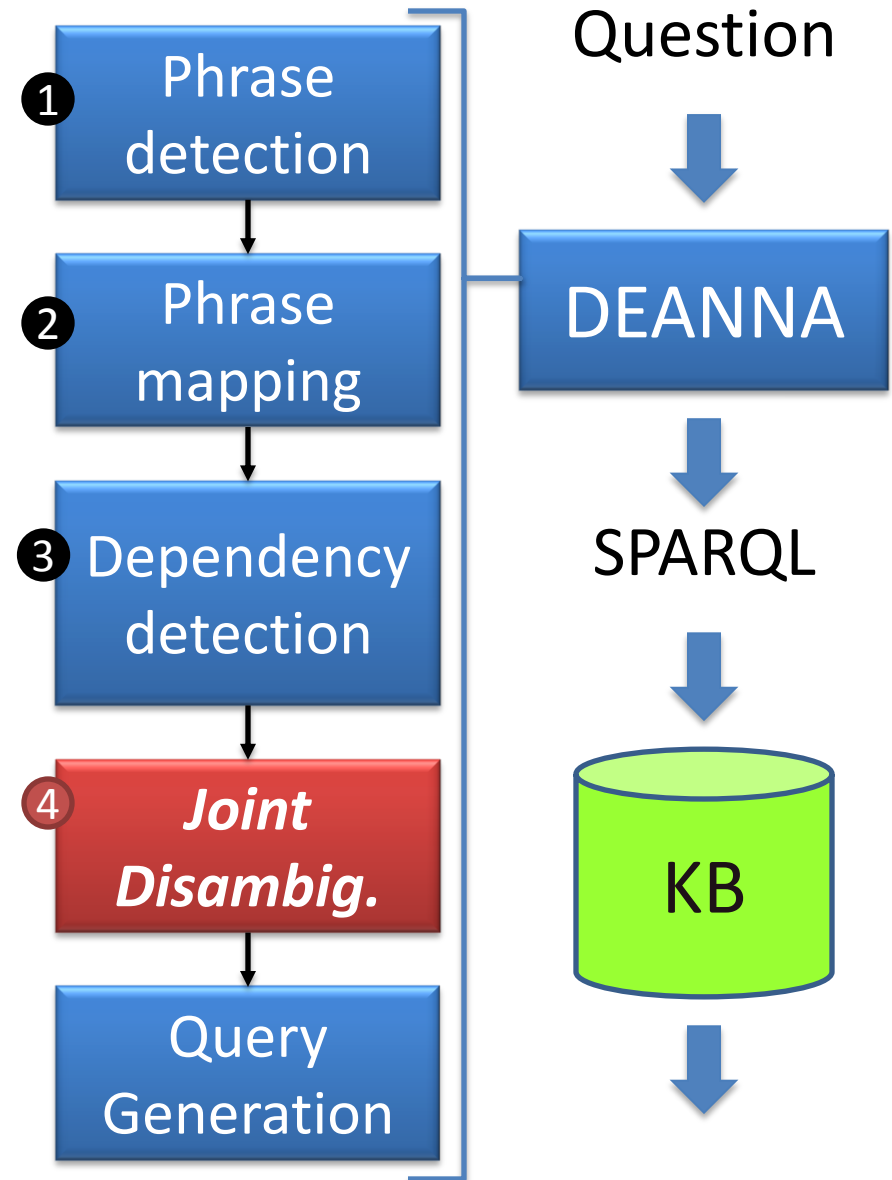
Classes translate to **type-constrained variables**

→ Every semantic triple should have a class to *join & project*!



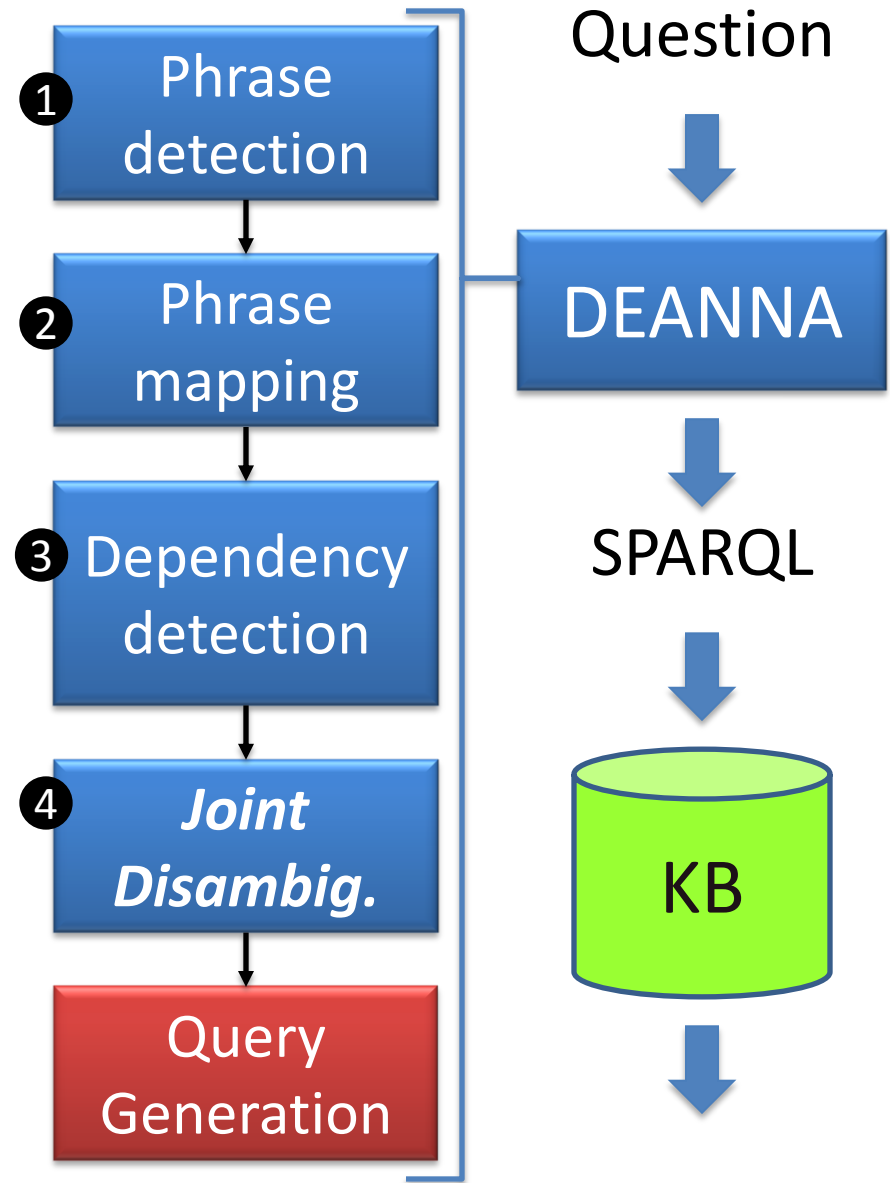
Outline

5 Experiments & Results

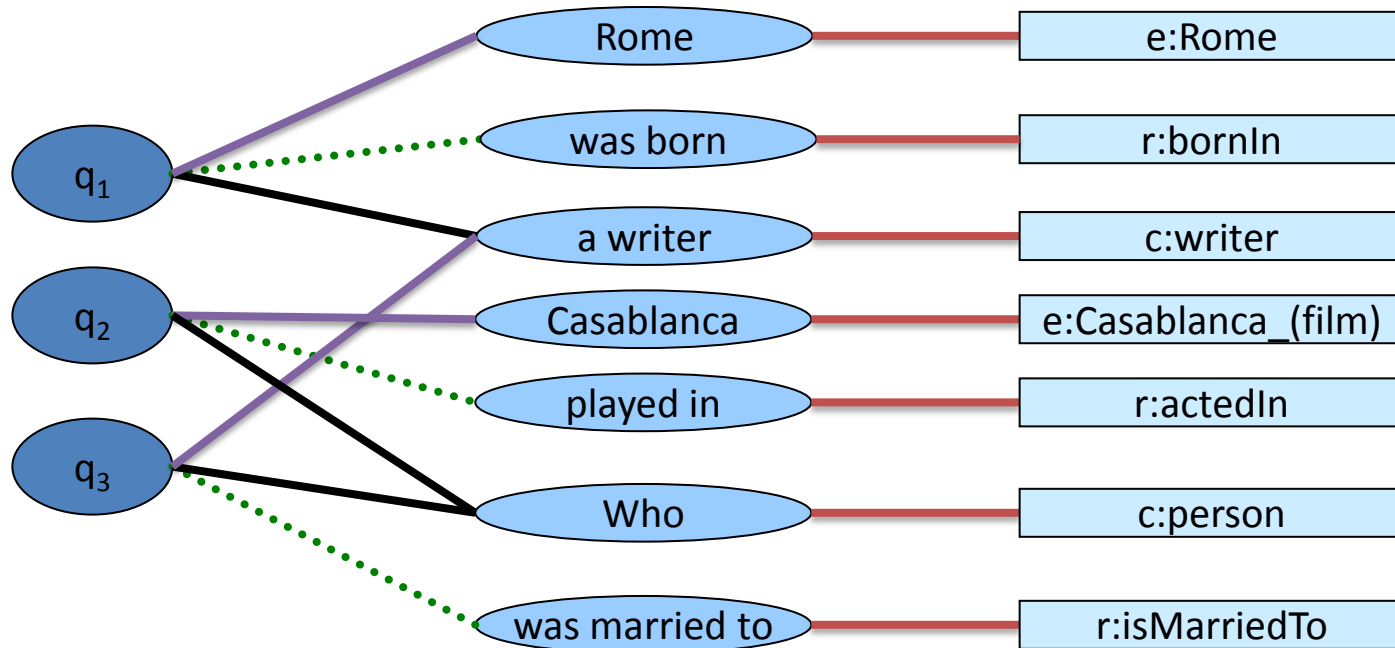


Outline

5 Experiments & Results



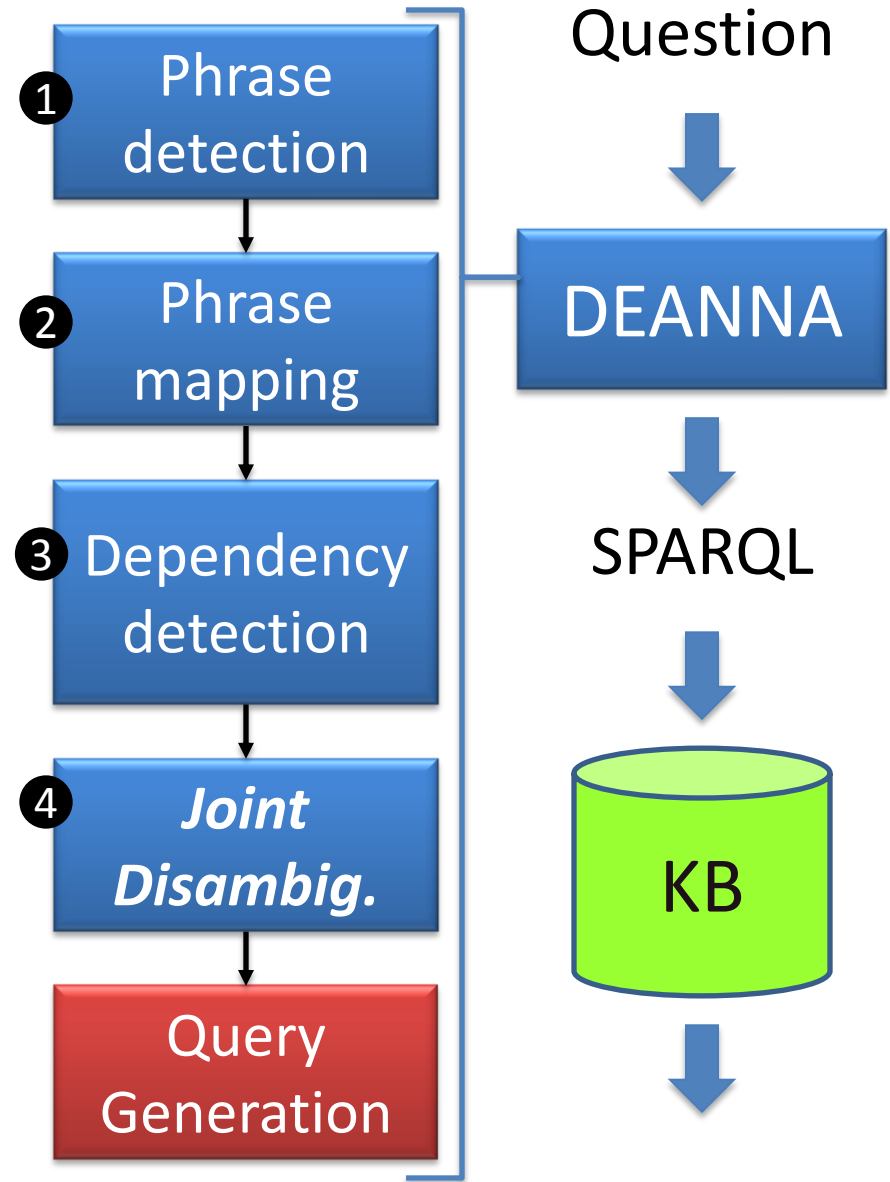
Structured Query Generation



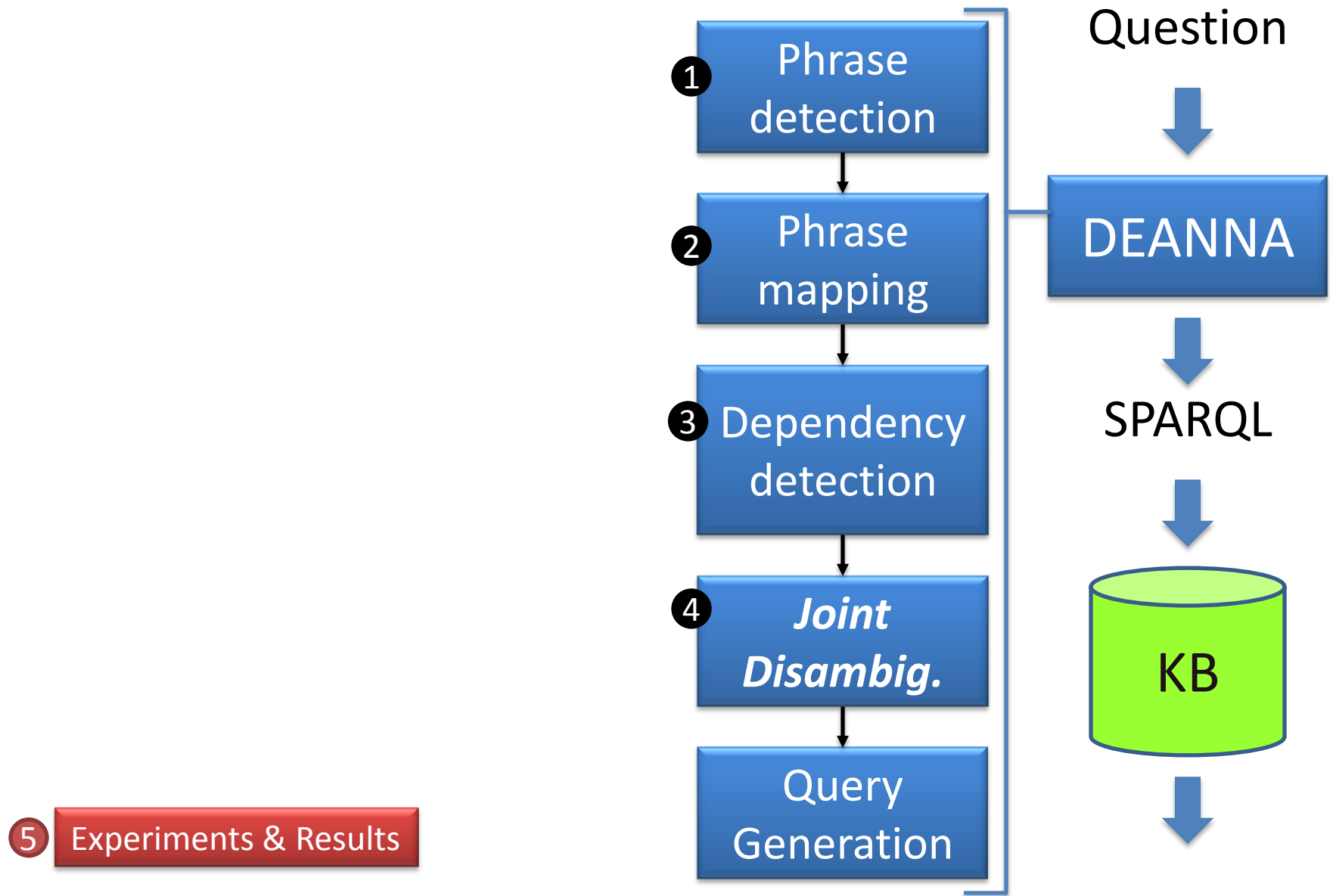
```
SELECT ?p WHERE {  
  ?w type writer .  
  ?w bornIn Rome .  
  ?p type person .  
  ?p actedIn Casablanca_(film) .  
  ?p isMarriedTo ?w }
```

Outline

5 Experiments & Results



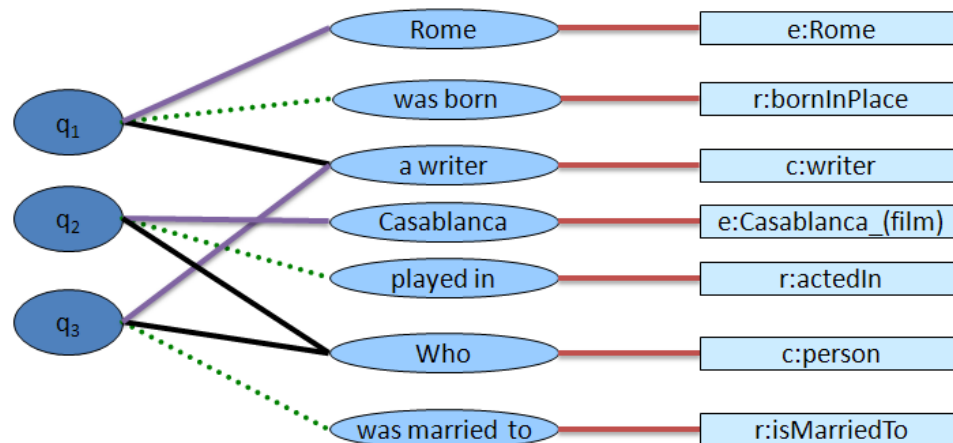
Outline



Evaluation Methodology

- **3-stage** evaluation for more insight:

1. Disambiguation
2. Query generation
3. Query answering

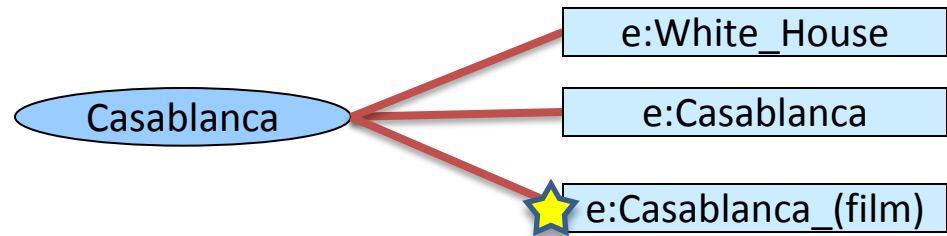


- We rely on **human judges**

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



Benchmark	QALD-1	NAGA
<i>COV</i> _{macro}	0.973	0.934
<i>prec</i> _{macro}	1.000	0.934
<i>COV</i> _{micro}	0.963	0.945
<i>prec</i> _{micro}	1.000	0.941

cov = correct/ideal

prec = correct / retrieved

Results 2/3: Query Generation

Issues:

- Incorrect disambiguation
- Incorrect dependencies.

Benchmark	QALD-1	NAGA
<i>COV_{macro}</i>	0.975	0.894
<i>prec_{macro}</i>	1.000	0.941
<i>COV_{micro}</i>	0.963	0.847
<i>prec_{micro}</i>	1.000	0.906

$cov = \#correct\ triples / \#ideal\ triples$

$prec = \#correct\ triples / \#retrieved\ triples$

Results 3/3: Answering

Benchmark	QALD-1	NAGA
<i>#questions</i>	27	44
<i>#queries</i>	20	41
<i>#satisfactory</i>	10	15
<i>#relaxed</i>	+3	+3

Results 3/3: Answering

Benchmark	QALD-1	NAGA
<i>#questions</i>	27	44
<i>#queries</i>	20	41
<i>#satisfactory</i>	10	15
<i>#relaxed</i>	+3	+3



unsatisfactory: main problem is
empty result set

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"]

Benchmark	QALD-1	NAGA
<i>#questions</i>	27	44
<i>#queries</i>	20	41
<i>#satisfactory</i>	10	15
<i>#relaxed</i>	+3	+3

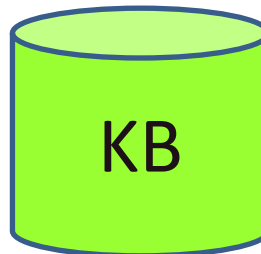
unsatisfactory: main problem is
empty result set

Wrap-up

Question



SPARQL



Answers

Wrap-up

Question

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



DEANNA



SPARQL



KB



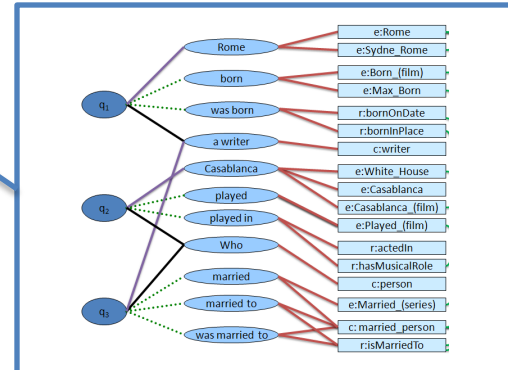
Answers

Wrap-up

Question

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

DEANNA



SPARQL

KB

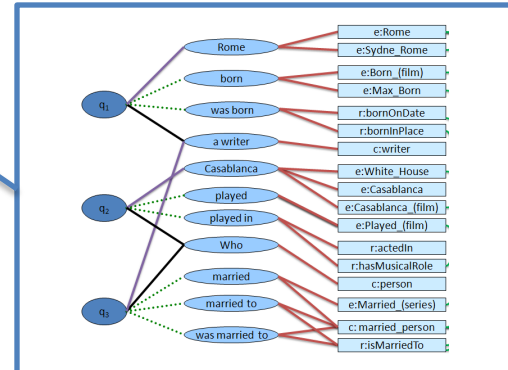
Answers

Wrap-up

Question

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

DEANNA



SPARQL

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

KB

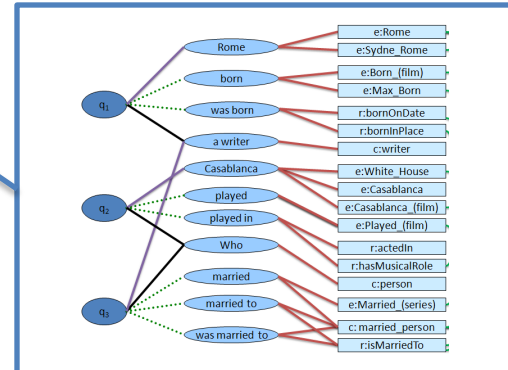
Answers

Wrap-up

Question

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

DEANNA

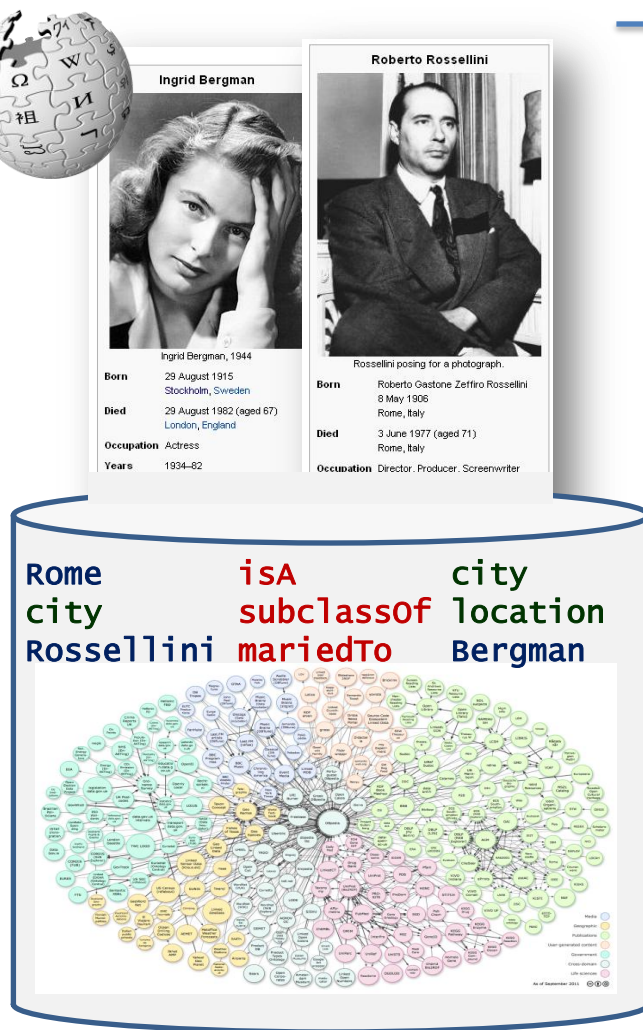


SPARQL

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

KB

Answers

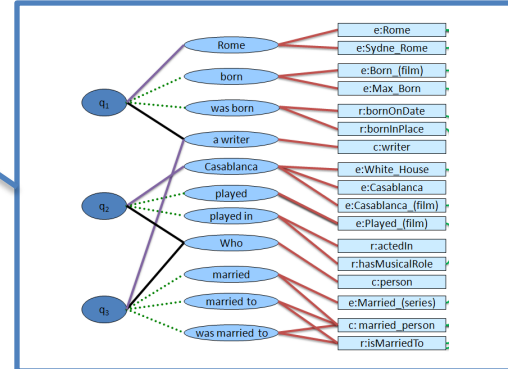


Wrap-up

Question

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

DEANNA

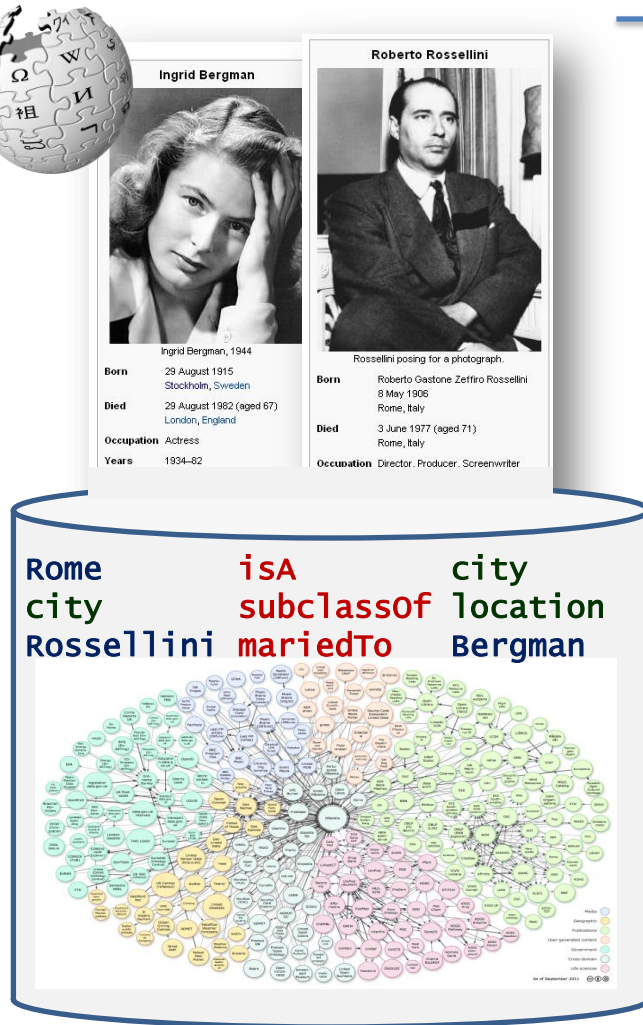


SPARQL

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

KB

Answers



Thank you.

QUESTIONS/COMMENTS?

bit.ly/mpi-deanna

