

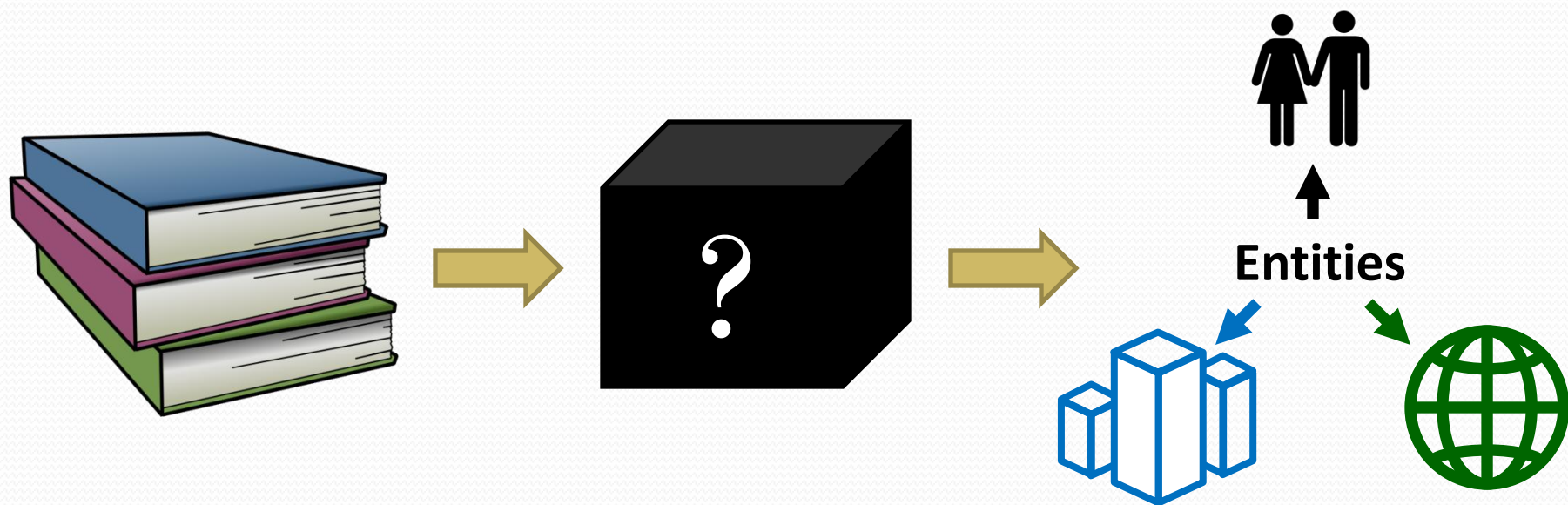
Entity Extraction:  
From Unstructured Text  
to  
DBpedia RDF Triples

Peter Exner  
Pierre Nugues



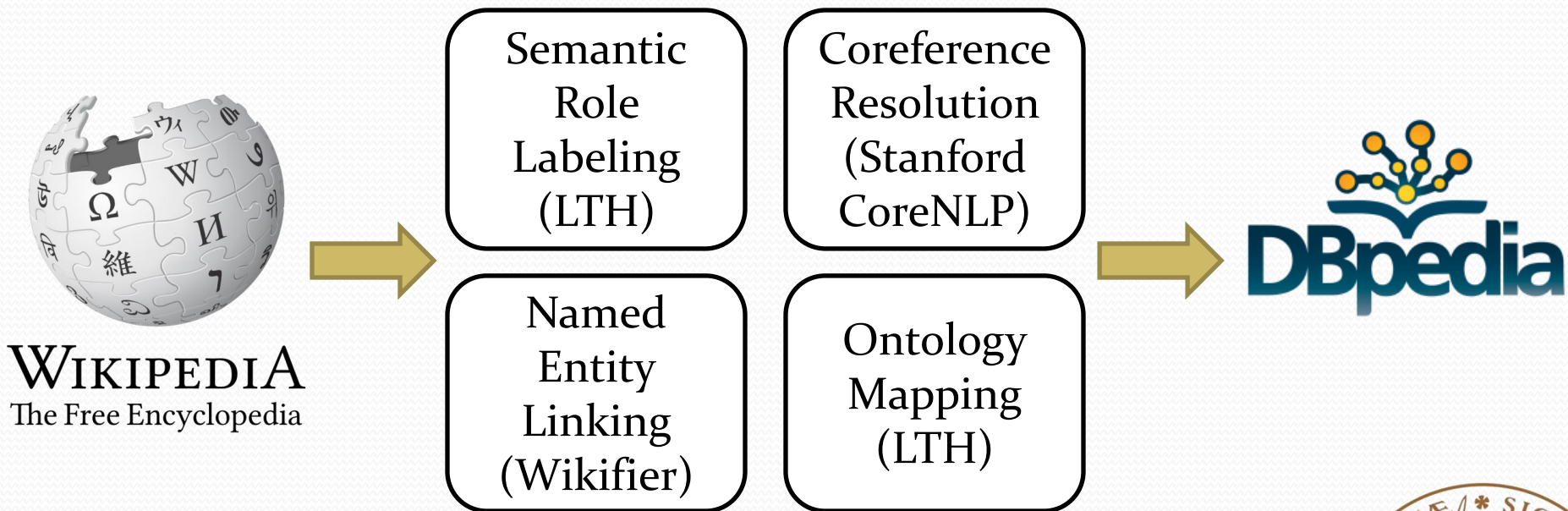
# Goal

- Extract entities from natural language.



# Approach

- Extend DBpedia
- Use Wikipedia as corpus



# Propbank

---

- Propbank
  - An annotated corpus of semantic roles: verbal propositions and their arguments, called **rolesets**.
  - Contains more than **7,000 rolesets**.

**A1**      **bear.02**      **AM-TMP**  
“**Luc Besson** (**born 18 March 1959**) is a French film director, writer, and producer.

**Predicate:** **bear.02**, *bear children*

**Roles:**

**Ao:** *mother*

**A1:** *child*

**Modifiers:**

**AM-TMP:** *temporal*

**AM-LOC:** *locative*



# Example – from text to DBpedia triple

## Wikipedia:

“Luc Besson (born 18 March 1959) is a French film director, writer, and producer.”



## DBpedia:

<dbpedia:Luc\_Besson>

<dbpedia-owl:birthDate>

”1959-03-18”



# Example step 1

“ Luc Besson (born 18 March 1959) is a French film director, writer, and producer. ”

“ Besson was born in Paris to parents who were both Club Med scuba diving instructors. ”



Semantic Role Labeling &  
Coreference Resolution

A1      bear.02      AM-TMP

“Luc Besson (born 18 March 1959) is a French film director, writer, and producer. ”

↑  
Coreference chain

A1      bear.02      AM-LOC

A2

“Besson was born in Paris to parents who were both Club Med scuba diving instructors.”

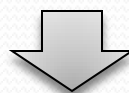


# Example step 2

**A1**      **bear.02**      **AM-TMP**  
“**Luc Besson** (**born 18 March 1959**) is a French film director, writer, and producer.”



**A1**      **bear.02**      **AM-LOC**      **A2**  
“**Besson** was **born in Paris** to parents who were both Club Med scuba diving instructors.”



Entity Extraction &  
Mention Entity Linking

URI from article  
<dbpedia:**Luc\_Besson**>

<propank:**bear.02.AM-TMP**>

”**1959-03-18**”

URI from Wikifier  
<dbpedia:**Besson**>

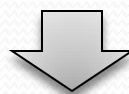
<propank:**bear.02.AM-LOC**>

URI from Wikifier  
<dbpedia:**Paris**>



# Example step 3

URI from article <dbpedia:Luc_Besson>	<propbank:bear.02.AM-TMP>	"1959-03-18"
URI from Wikifier <dbpedia:Besson>	<propbank:bear.02.AM-LOC>	URI from Wikifier <dbpedia:Paris>



Coreference Inference &  
Ontology Mapping

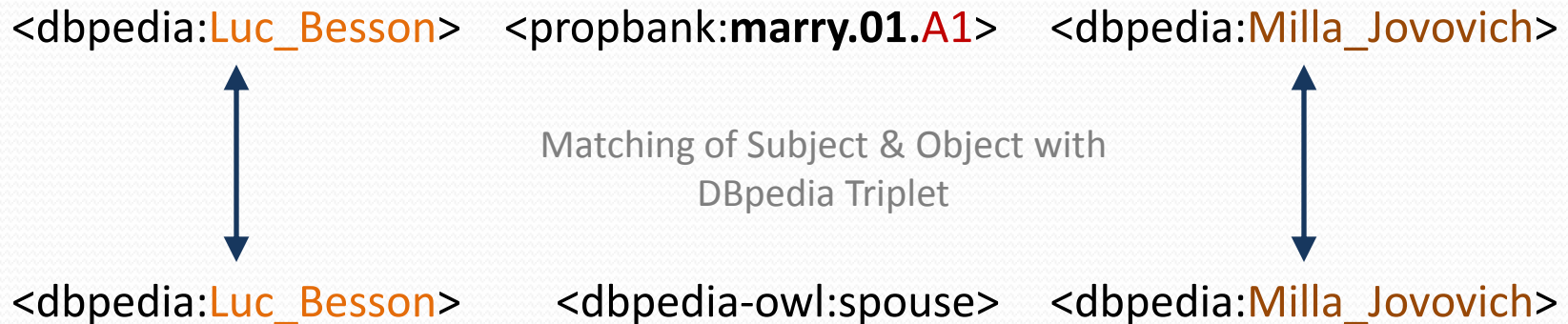
<dbpedia:Luc_Besson>	<dbpedia-owl:birthDate>	"1959-03-18"
↑ Coreference scoring module		
<dbpedia:Luc_Besson>	<dbpedia-owl:birthPlace>	<dbpedia:Paris>



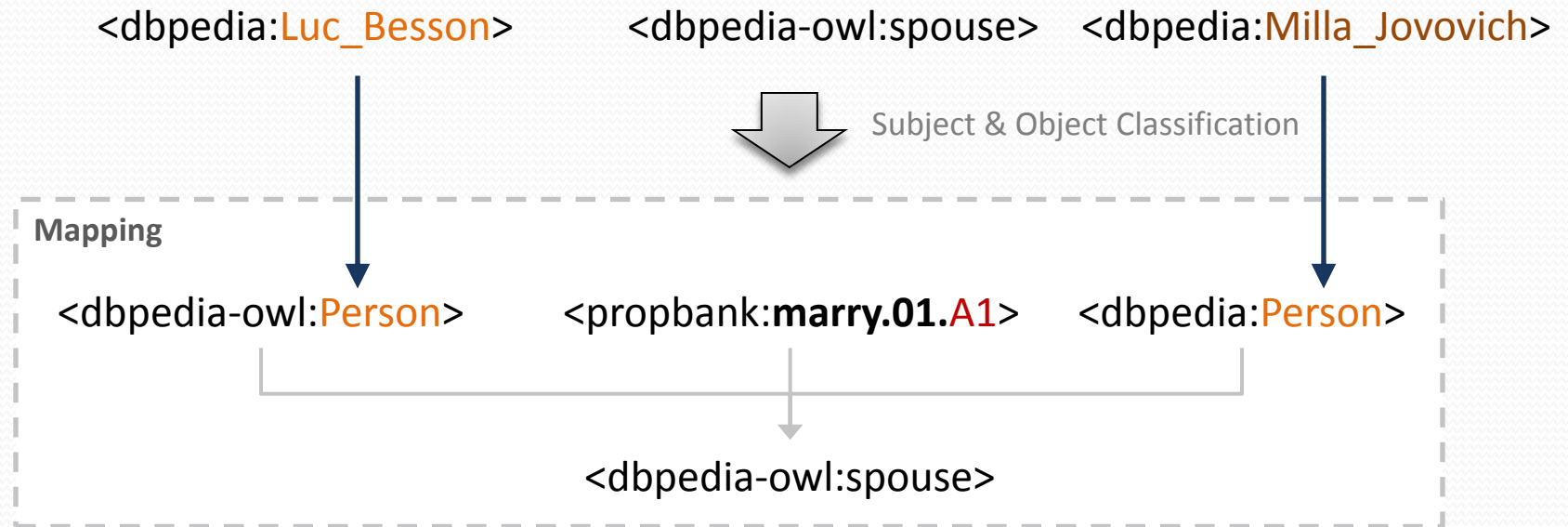


# Ontology Mapping – learning step 1

“ Besson married Milla Jovovich on 14 December 1997. ”



# Ontology Mapping – learning step 2



# Ontology Mapping – reusing step 1

“ On April 30, 2008, Carey married Cannon at her private estate...”



Triple extraction

<dbpedia:Mariah\_Carey>

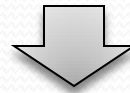
<propbank:marry.01.A1>

<dbpedia:Nick\_Cannon>



# Ontology Mapping – reusing step 2

“ On April 30, 2008, Carey married Cannon at her private estate...”

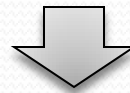


Triple extraction

<dbpedia:Mariah\_Carey>

<propbank:marry.01.A1>

<dbpedia:Nick\_Cannon>



Subject & Object Generalization

<dbpedia-owl:Person>

<propbank:marry.01.A1>

<dbpedia-owl:Person>



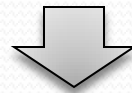
# Ontology Mapping – reusing step 3

“ On April 30, 2008, Carey married Cannon at her private estate... ”



Triple extraction

<dbpedia:Mariah\_Carey>    <propbank:marry.01.A1>    <dbpedia:Nick\_Cannon>



Subject & Object Generalization

<dbpedia-owl:Person>    <propbank:marry.01.A1>    <dbpedia-owl:Person>



Mapping

<dbpedia:Mariah\_Carey>    <dbpedia-owl:spouse>    <dbpedia:Nick\_Cannon>



# Mappings learned

- In total 322 mappings

Subject	Predicate	Object	DBpedia
Person	bear.02.AM-LOC	Place	birthPlace
Person	bear.02.AM-TMP	Date	birthDate
Person	marry.01.A1	Person	spouse
Place	locate.01.AM-LOC	Place	isPartOf
Place	open.01.AM-TMP	Date	openingDate
Place	build.01.AM-TMP	Numeric	yearOfConstruction
Organization	serve.01.A2	Place	broadcastArea
Organization	own.01.A1	Organization	subsidiary
Organization	include.01.A2	Person	bandMember



# Results - Triples

Type	Count
Total (from 114,895 articles)	1,023,316
DBpedia Mapped Triples	189,610
...of which 15,067 already exist in DBpedia	15,067
Unmapped Triples	833,706

- 833,706 triples couldn't be mapped due to lacking a triple to learn from in Dbpedia, e.g.:

<dbpedia:lan\_Holm>

<propbank:portray.01>

<dbpedia:Hercule\_Poirot>

... interlinking with other datasets (LinkedMDB etc.) is needed!



# Precision / Recall

Precision	Recall	F1
74.3	59.9	66.3

Error source	
Subject	27.2
Predicate	46.4
Object	26.4





# Datasets available

---

An archive of entities in N-Triple format is available at:

<http://semantica.cs.lth.se>



# Q&A

---

