



max planck institut
informatik

YAGO3:

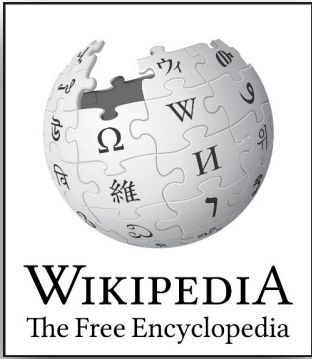
A Knowledge Base from Multilingual Wikipedias

Farzaneh Mahdisoltani

Joanna Biega

Fabian M. Suchanek

CIDR 2015



John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see [Coltrane \(disambiguation\)](#).
John William Coltrane, also known as "**Trane**" (September 23, 1926 – , **saxophonist** and composer. Working in the **bebop** and **hard bop** idioms e

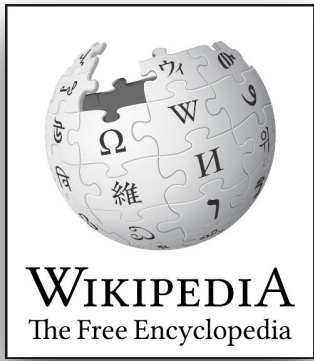
Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde](#)

John Coltrane

Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York



John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see [Coltrane \(disambiguation\)](#).
John William Coltrane, also known as "**Trane**" (September 23, 1926 – ,
[saxophonist](#) and composer. Working in the [bebop](#) and [hard bop](#) idioms e

Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) |
| [American jazz bandleaders](#) | [American jazz composers](#)
| [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde](#)

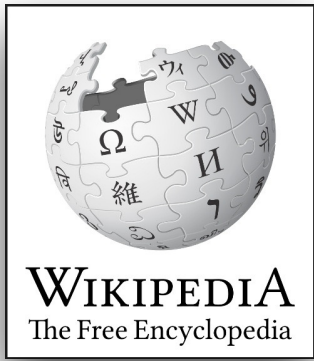
John Coltrane

Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York





John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see Coltrane (disambiguation).
John William Coltrane, also known as **"Trane"** (September 23, 1926 – July 17, 1967), was an American **saxophonist** and composer. Working in the **bebop** and **hard bop** idioms, he was one of the most influential and innovative jazz musicians of the 20th century.

Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde jazz](#)

John Coltrane



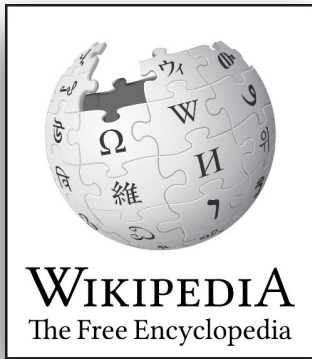
Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York



John_Coltrane



John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see *Coltrane (disambiguation)*.
John William Coltrane, also known as "**Trane**" (September 23, 1926 – July 17, 1967), was an American **saxophonist** and composer. Working in the **bebop** and **hard bop** idioms, he was one of the most influential and innovative jazz musicians of the 20th century.

Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde jazz](#)

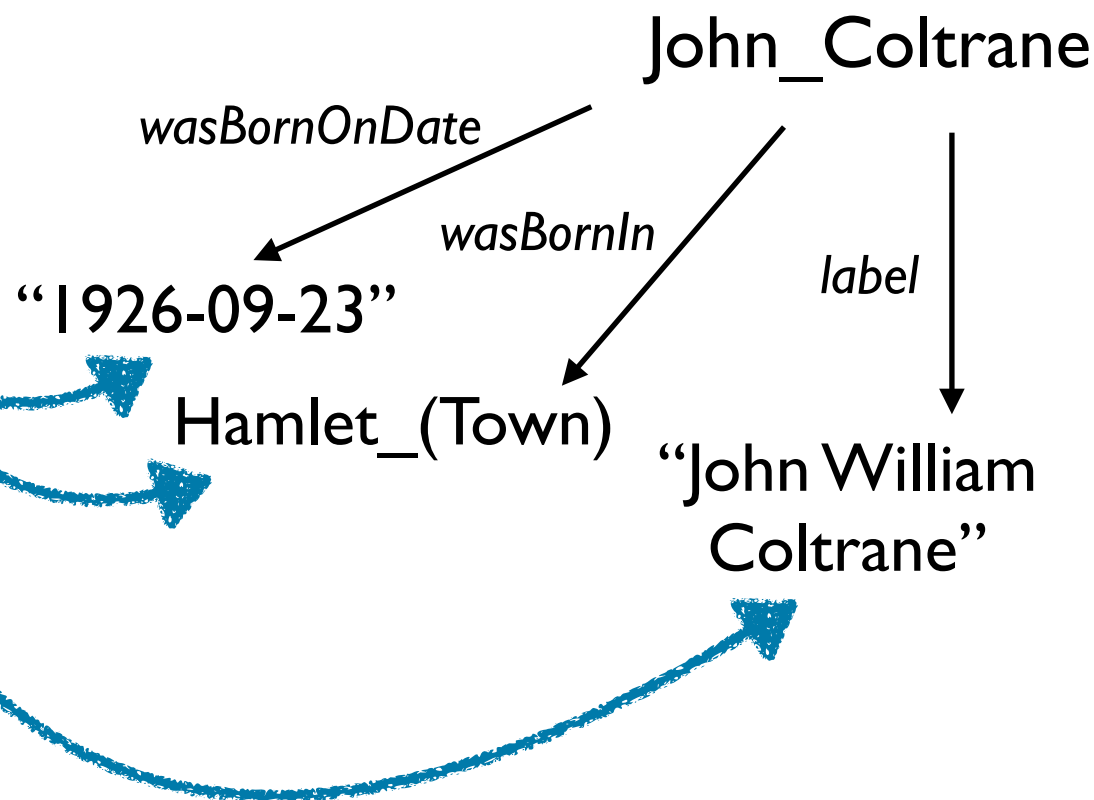
John Coltrane

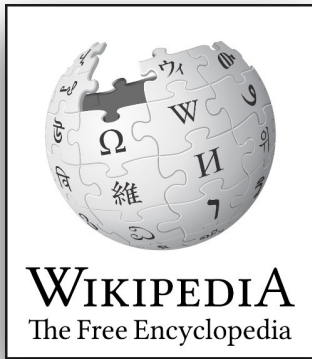


Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York





John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see *Coltrane (disambiguation)*.
John William Coltrane, also known as "**Trane**" (September 23, 1926 – July 17, 1967), was an American **saxophonist** and composer. Working in the **bebop** and **hard bop** idioms, he was one of the most influential and innovative jazz musicians of the 20th century.

Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde jazz](#)

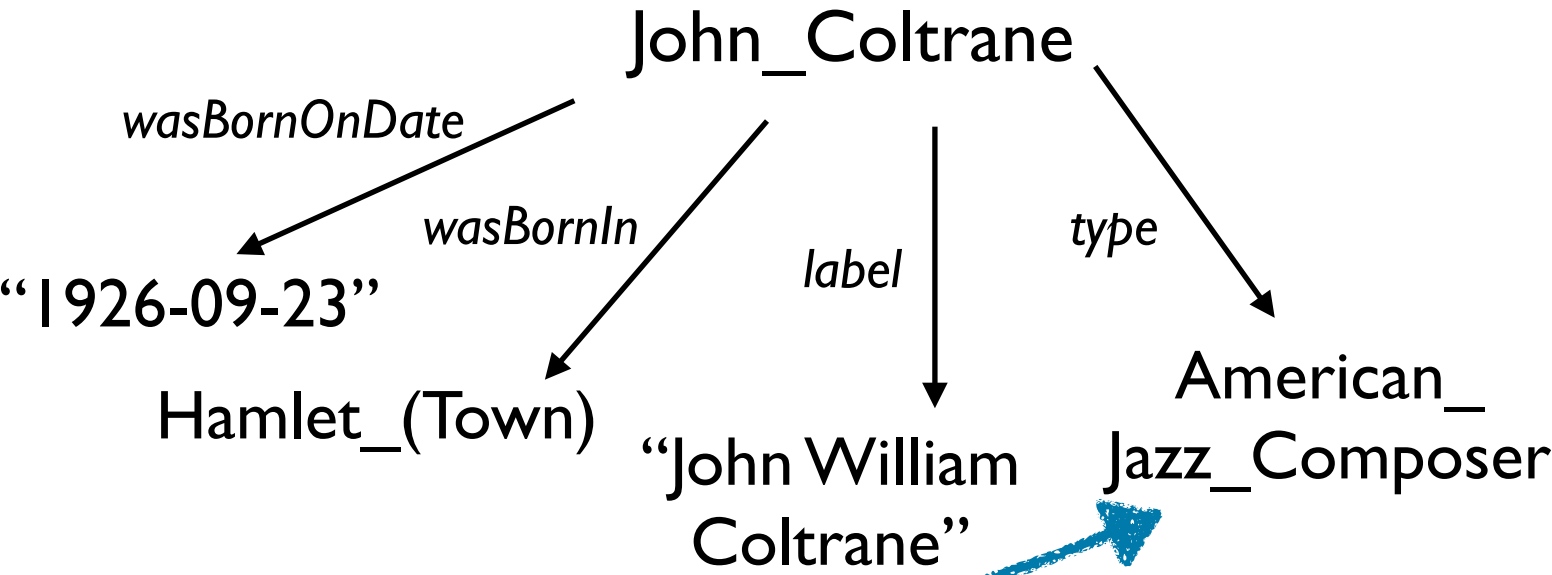
John Coltrane

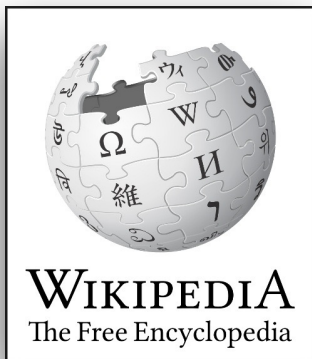


Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York





John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see *Coltrane (disambiguation)*.
John William Coltrane, also known as "**Trane**" (September 23, 1926 – July 17, 1967), was an American **saxophonist** and composer. Working in the **bebop** and **hard bop** idioms, he was one of the most influential and innovative jazz musicians of the 20th century.

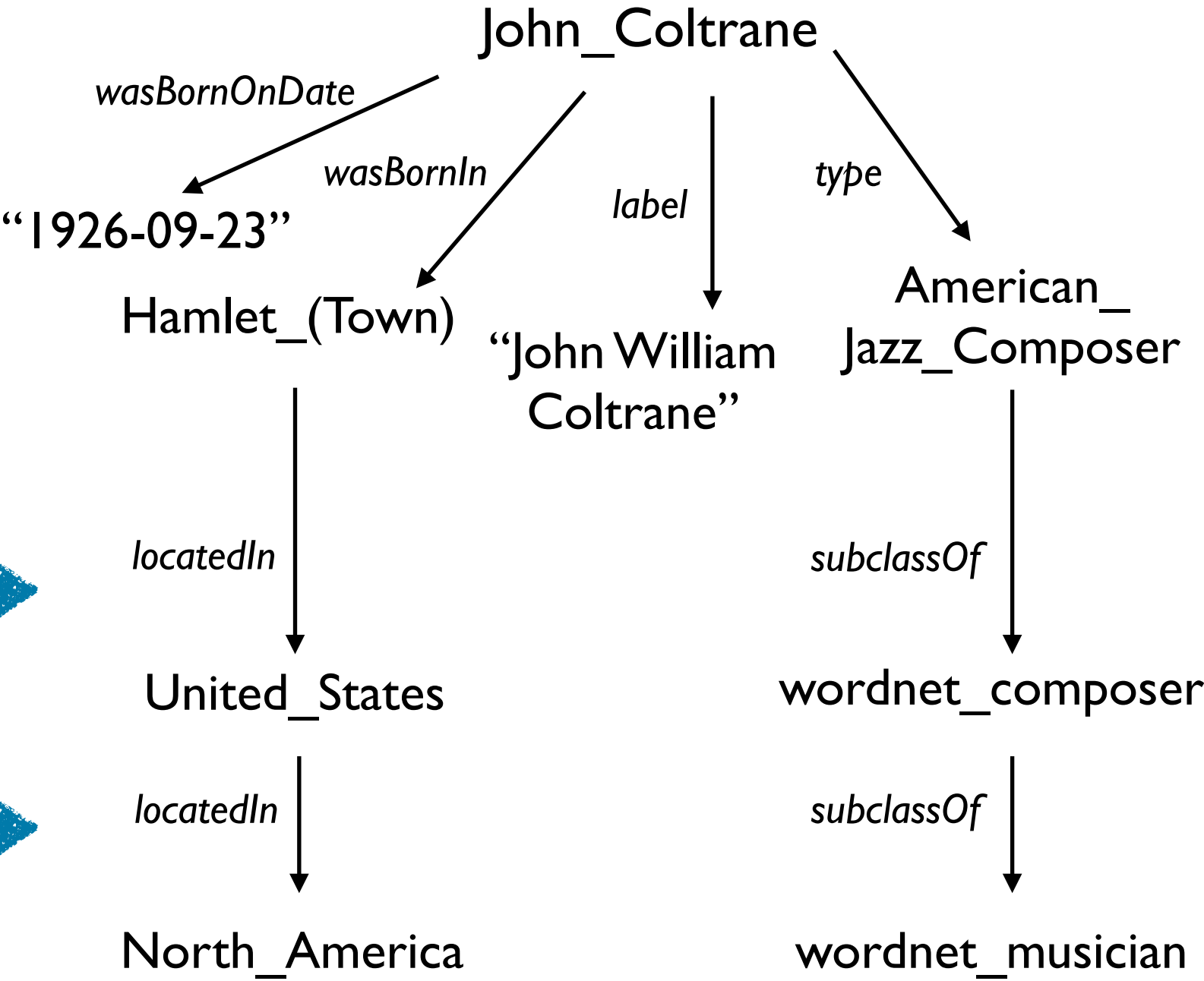
Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde](#)

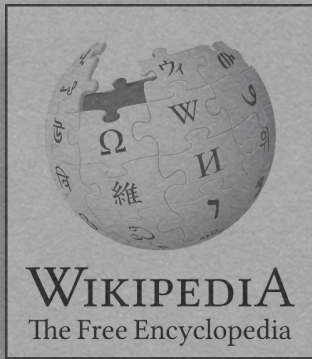
John Coltrane

Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina, USA
Died	July 17, 1967 (aged 40) Huntington, New York





John Coltrane

From Wikipedia, the free encyclopedia

"Coltrane" redirects here. For other uses, see *Coltrane (disambiguation)*.
John William Coltrane, also known as "**Trane**" (September 23, 1926 – July 17, 1967), was an American jazz saxophonist and composer. Working in the bebop and hard bop idioms early in his career, Coltrane later explored more avant-garde styles.

Categories: [John Coltrane](#) | [1926 births](#) | [1967 deaths](#) | [American jazz bandleaders](#) | [American jazz composers](#) | [Anglican saints](#) | [Atlantic Records artists](#) | [Avant-garde](#)

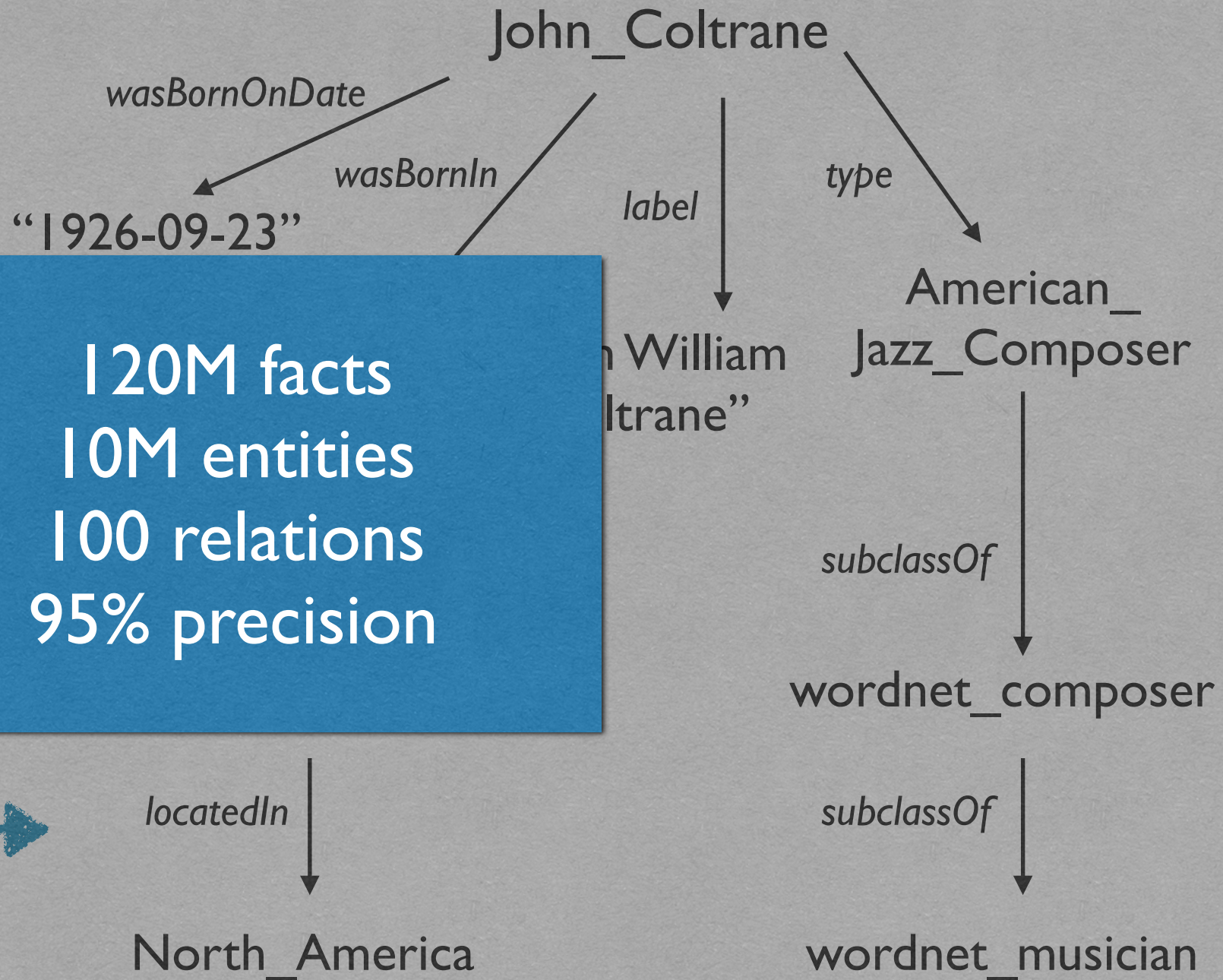
John Coltrane



Coltrane in 1963

Background information

Birth name	John William Coltrane
Also known as	"Trane"
Born	September 23, 1926 Hamlet, North Carolina
Died	July 17, 1967 (aged 40) Huntington, New York



GeoNames

120M facts
10M entities
100 relations
95% precision

WordNet
A lexical database for English



YAGO can be used in many ways

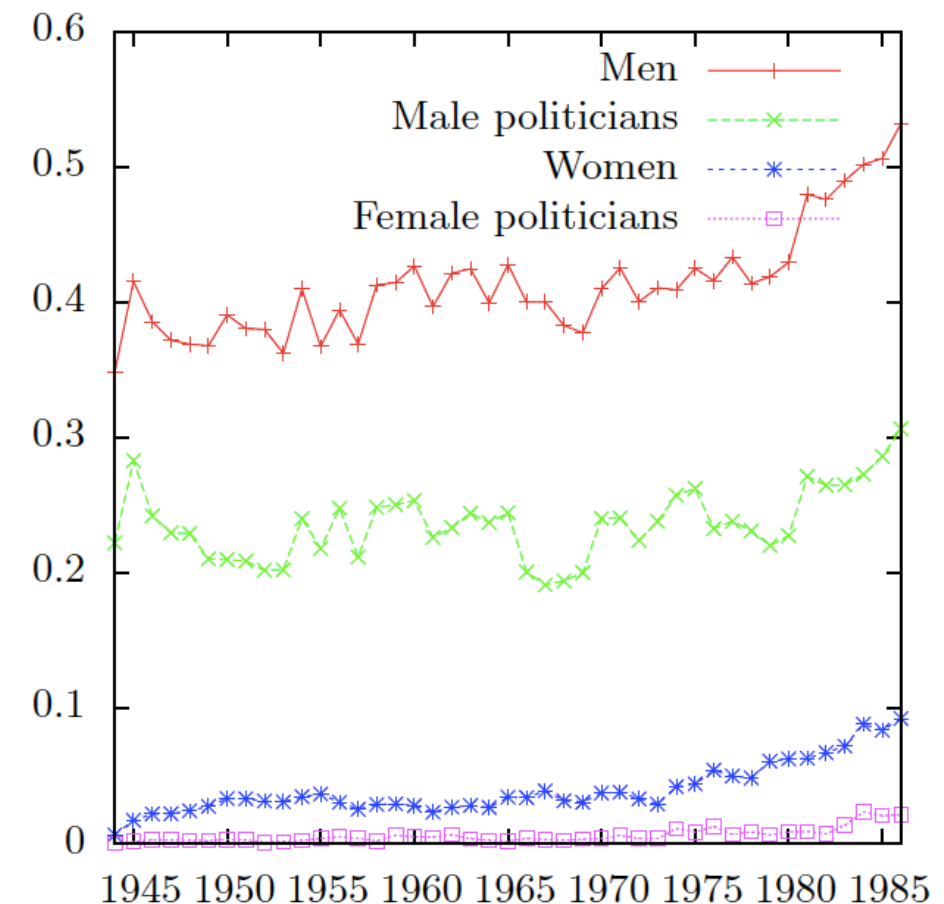


Named Entity
Disambiguation

YAGO can be used in many ways



Named Entity
Disambiguation



Semantic Culturomics

T. Huet, J. Biega, F. M. Suchanek, Mining History with Le Monde, AKBC2013

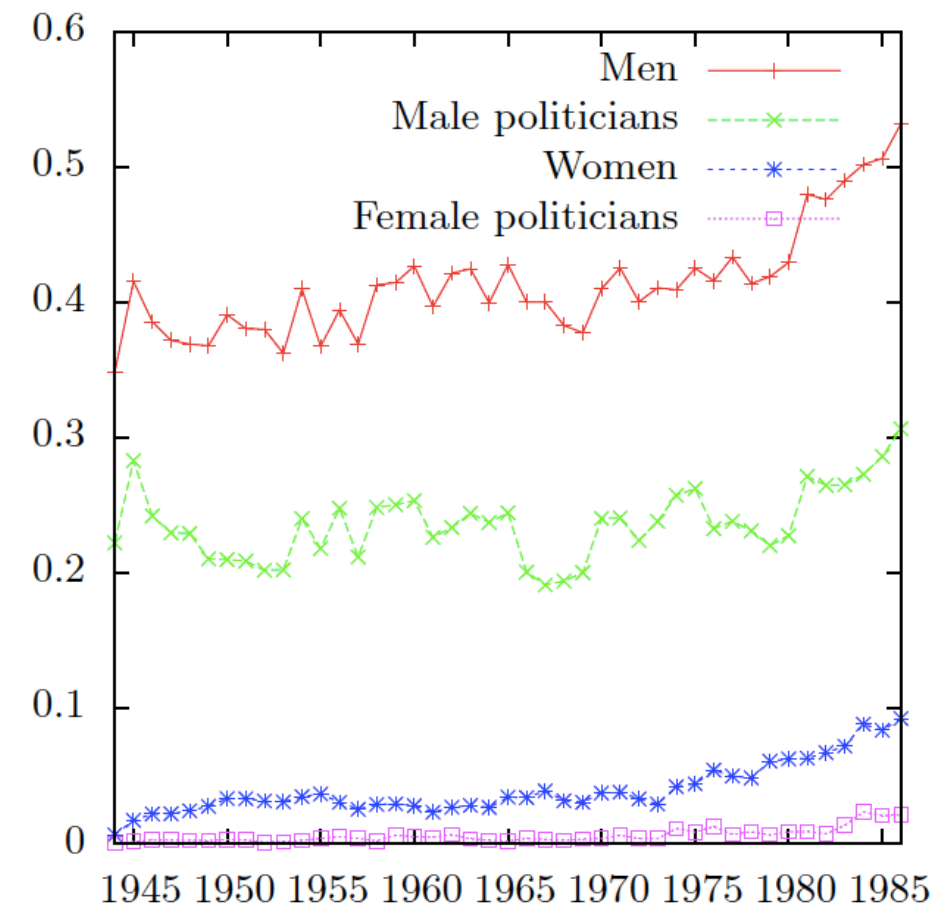
F. M. Suchanek, N. Preda, Semantic Culturomics, VLDB2014

J. Hoffart et al., Robust Disambiguation of Named Entities in Text, EMNLP2011

YAGO can be used in many ways



Named Entity
Disambiguation



Semantic Culturomics

Extending YAGO coverage would yield better results!

T. Huet, J. Biega, F. M. Suchanek, Mining History with Le Monde, AKBC2013

F. M. Suchanek, N. Preda, Semantic Culturomics, VLDB2014

J. Hoffart et al., Robust Disambiguation of Named Entities in Text, EMNLP2011

Multilingual wikipedias



Multilingual wikipedias



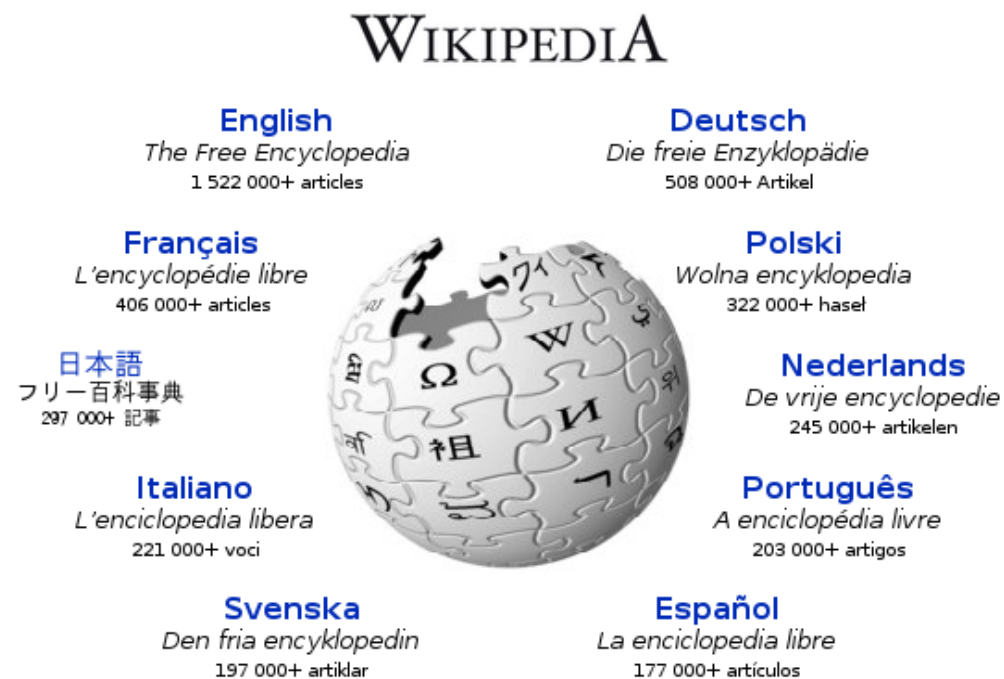
 Izabella_Olszewska



 Tadeusz_Jurasz



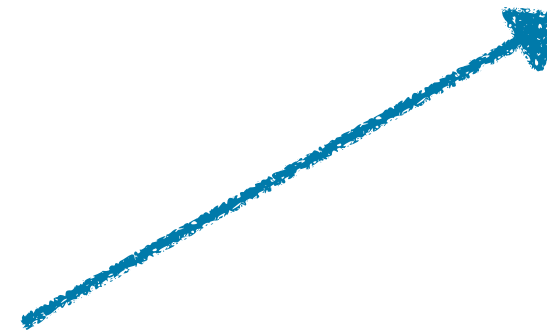
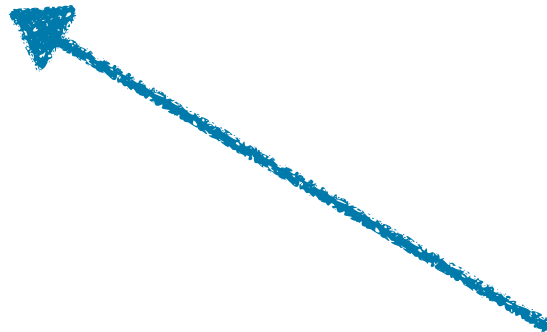
Multilingual wikipedias



Local facts

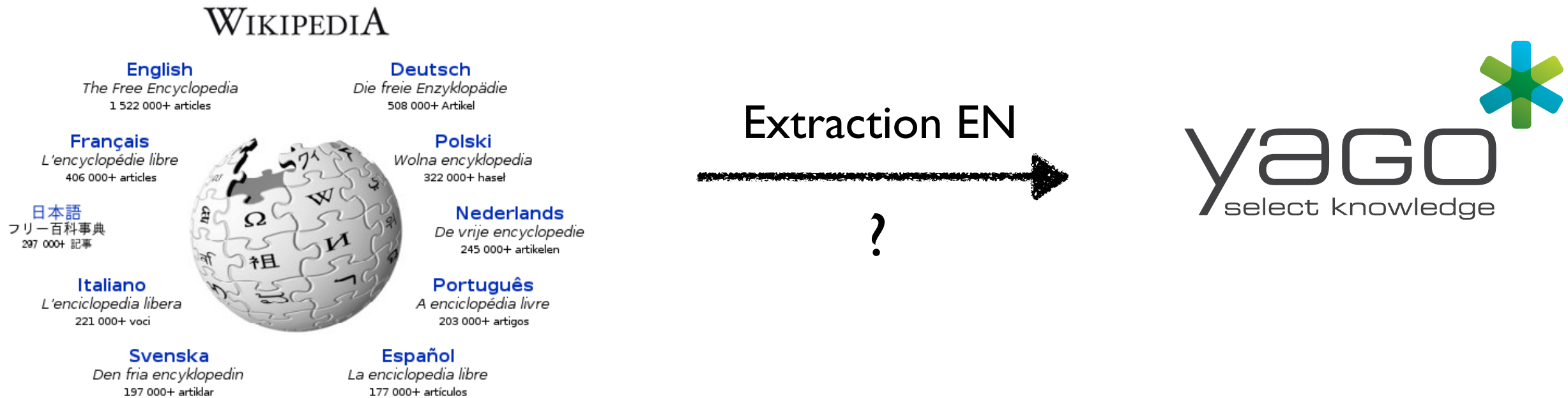


 Izabella_Olszewska isMarriedTo  Tadeusz_Jurasz

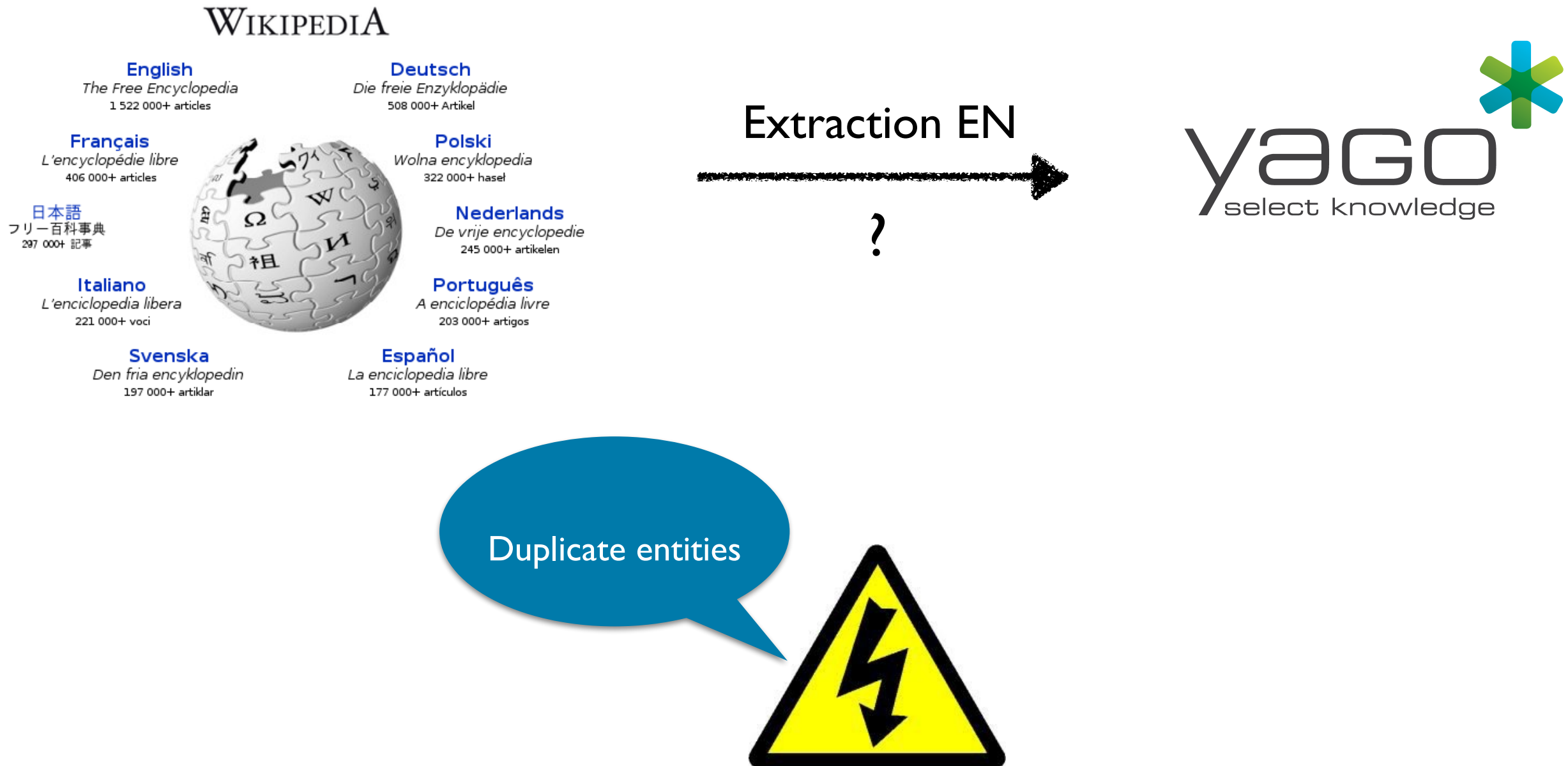


Local entities

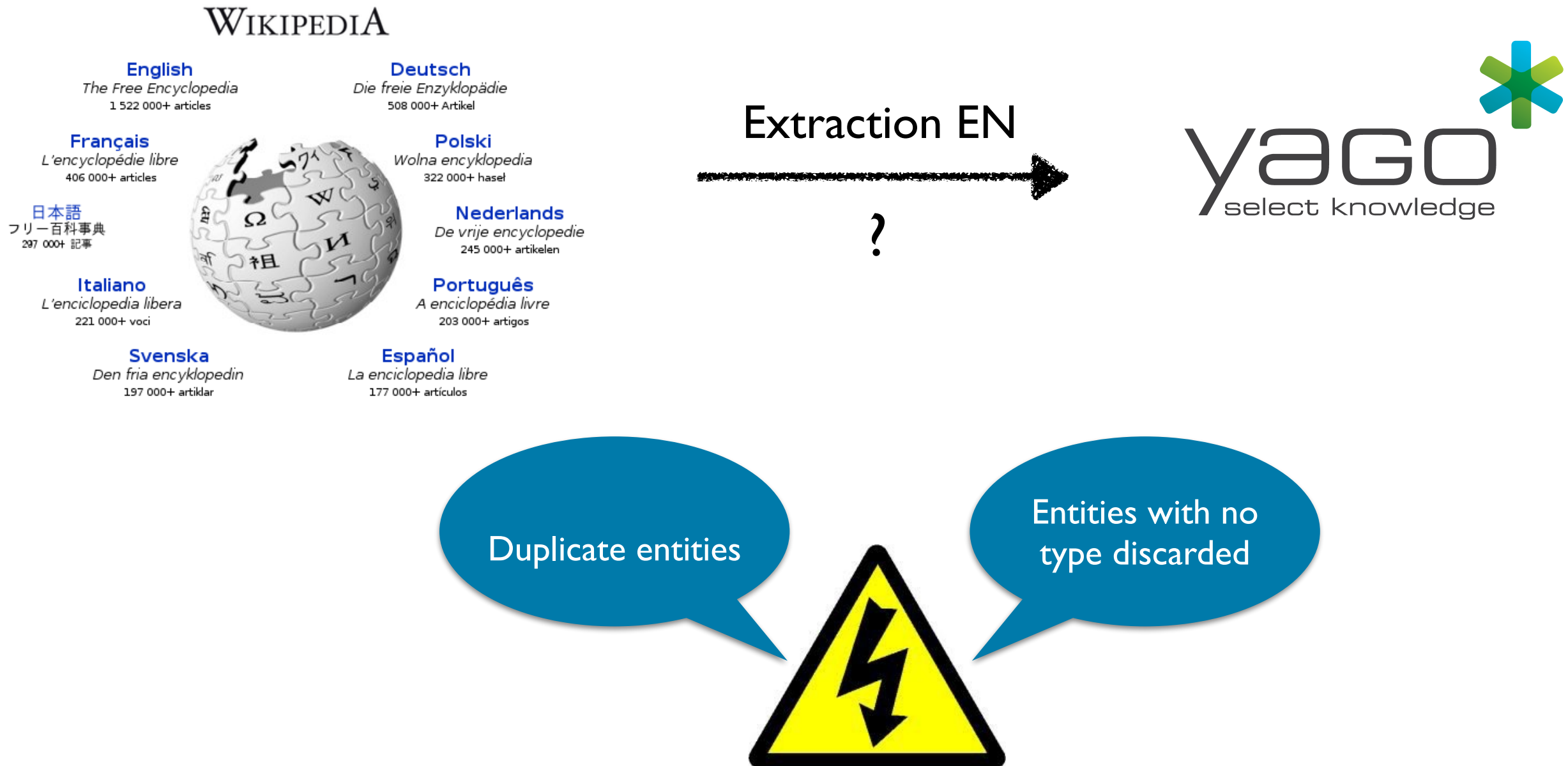
Running YAGO on multilingual wikipedias



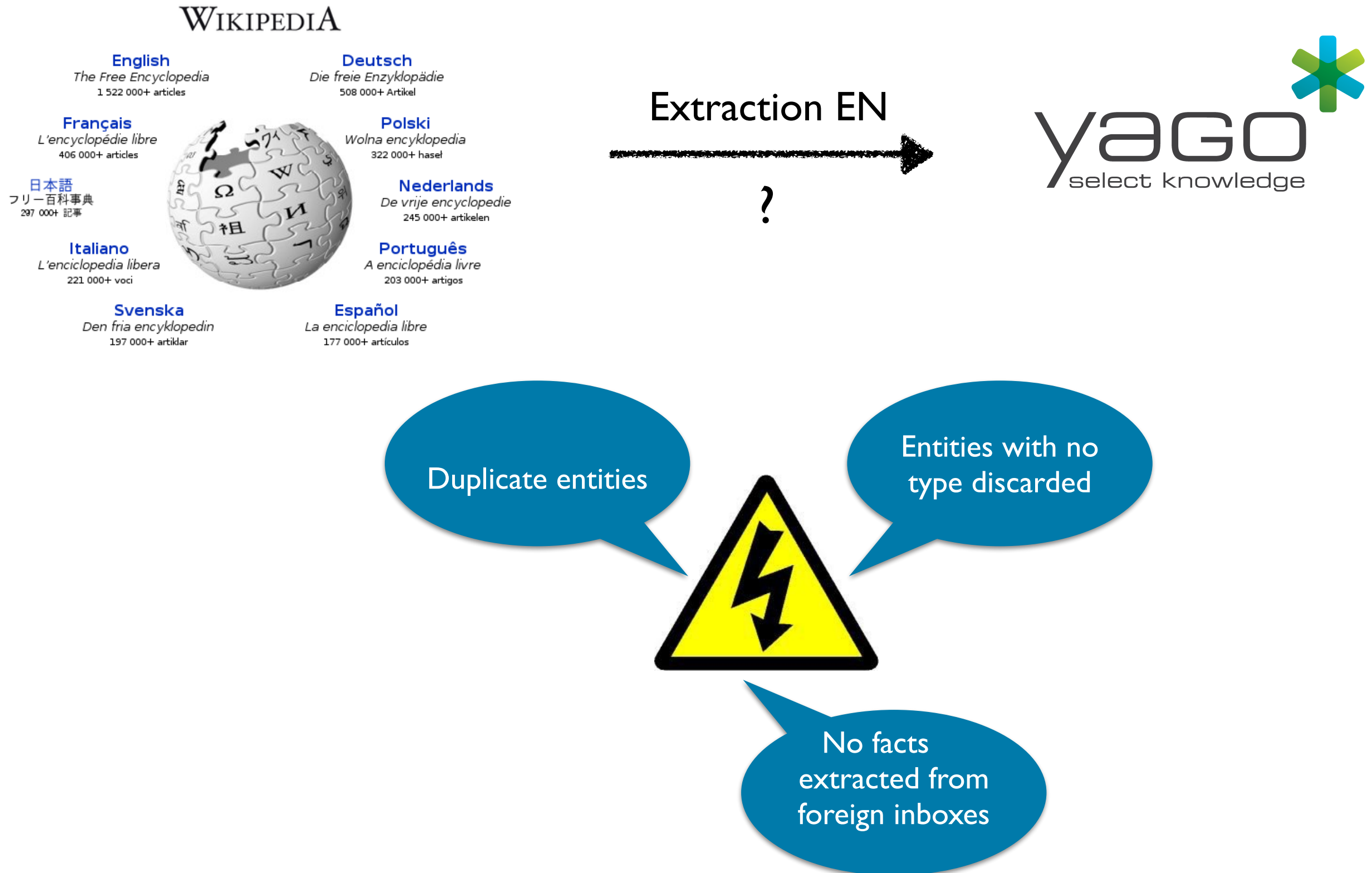
Running YAGO on multilingual wikipedias



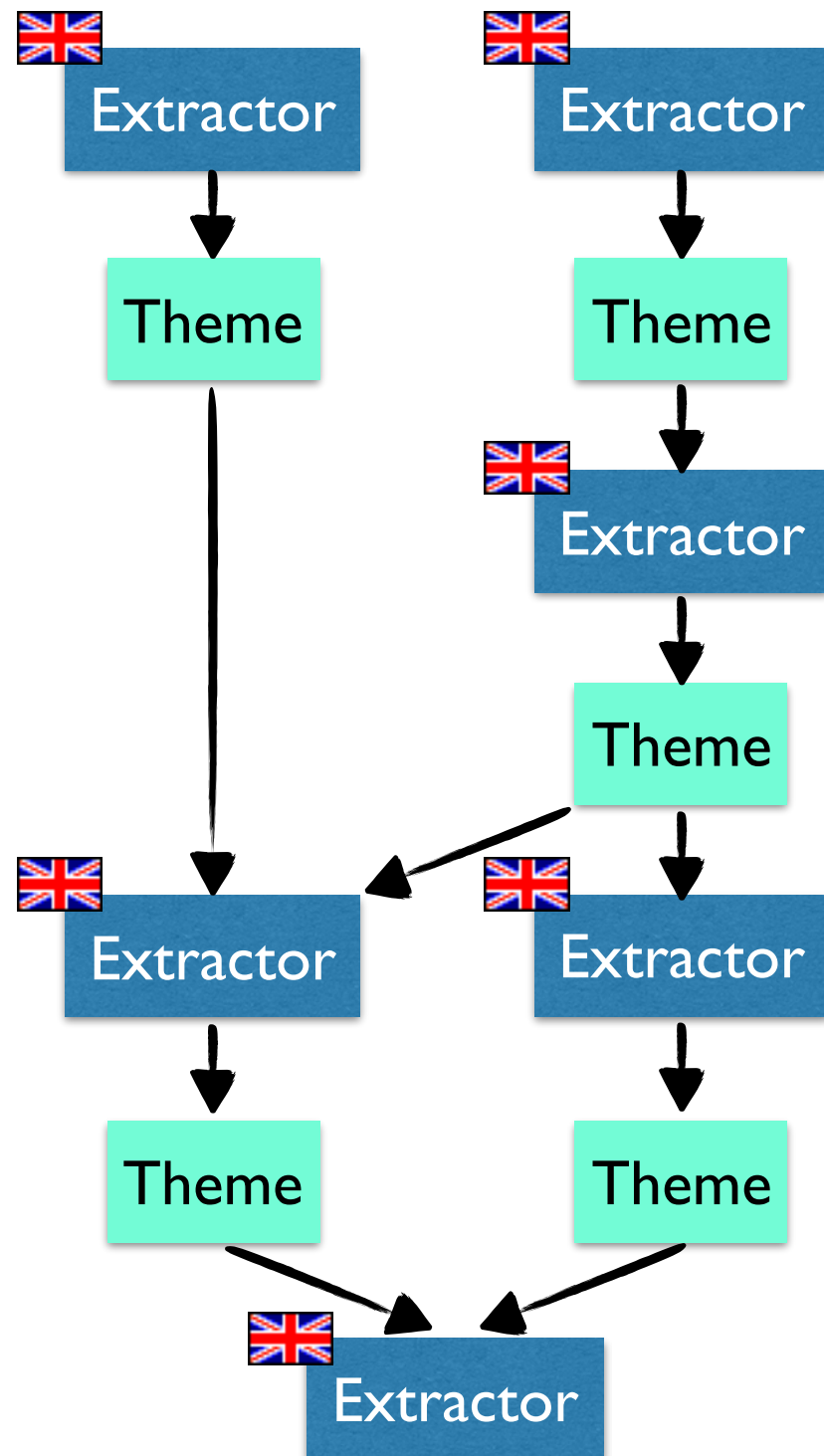
Running YAGO on multilingual wikipedias



Running YAGO on multilingual wikipedias

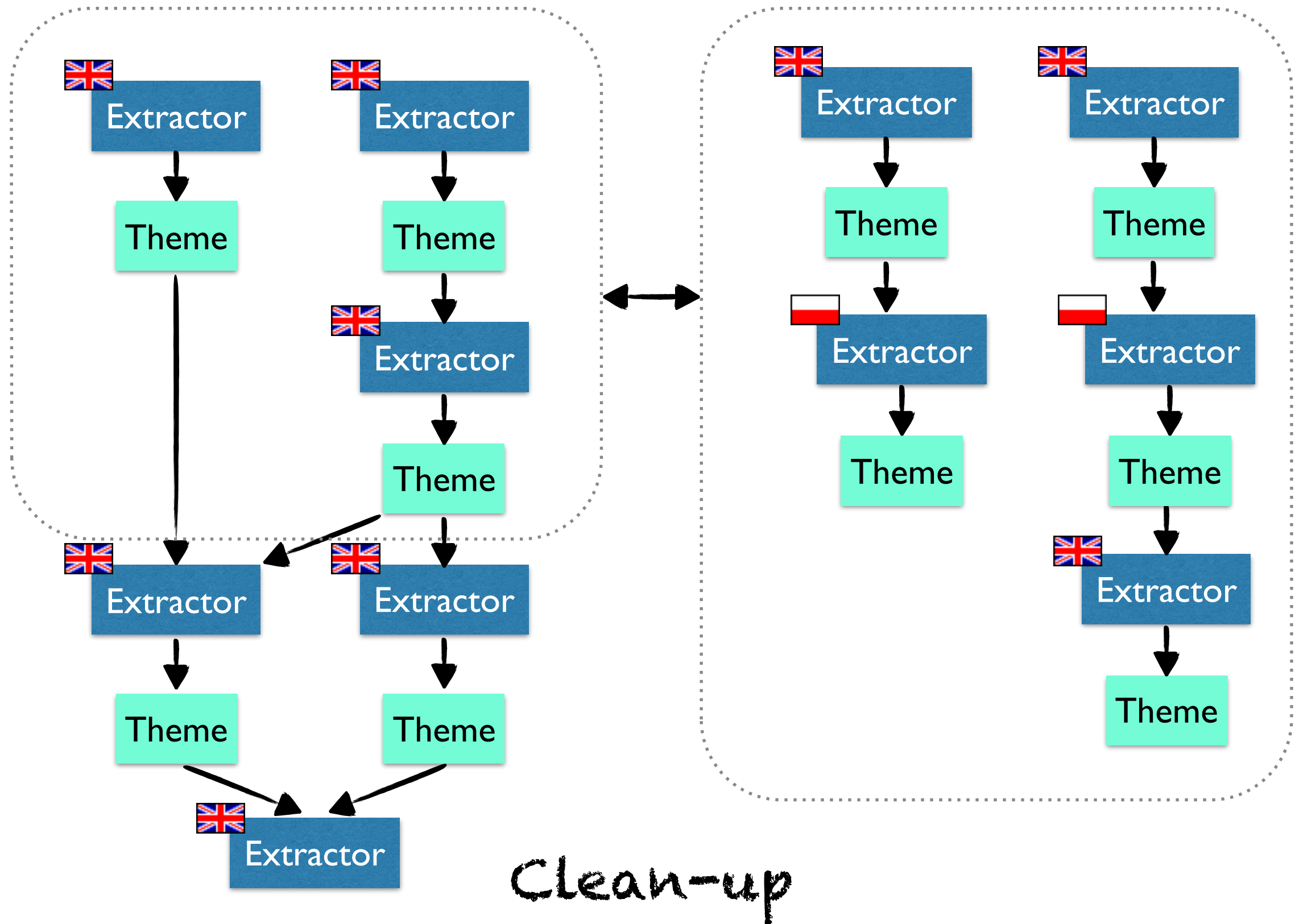


Running YAGO on multilingual wikipedias



Running YAGO on multilingual wikipedias

Raw extraction



Tasks

1. Entities
2. Types
3. Facts

I. Set of Entities



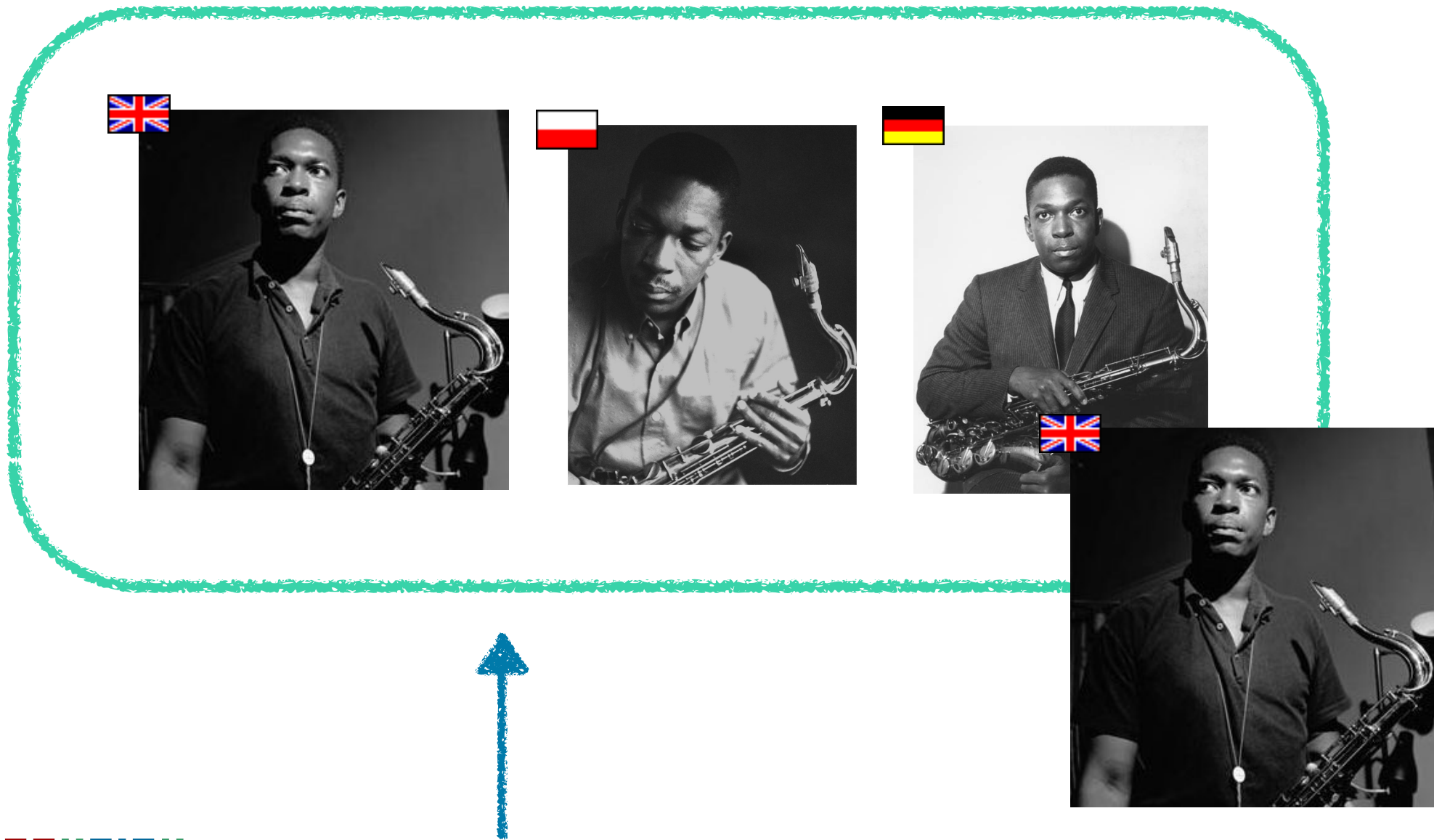
=?



=?

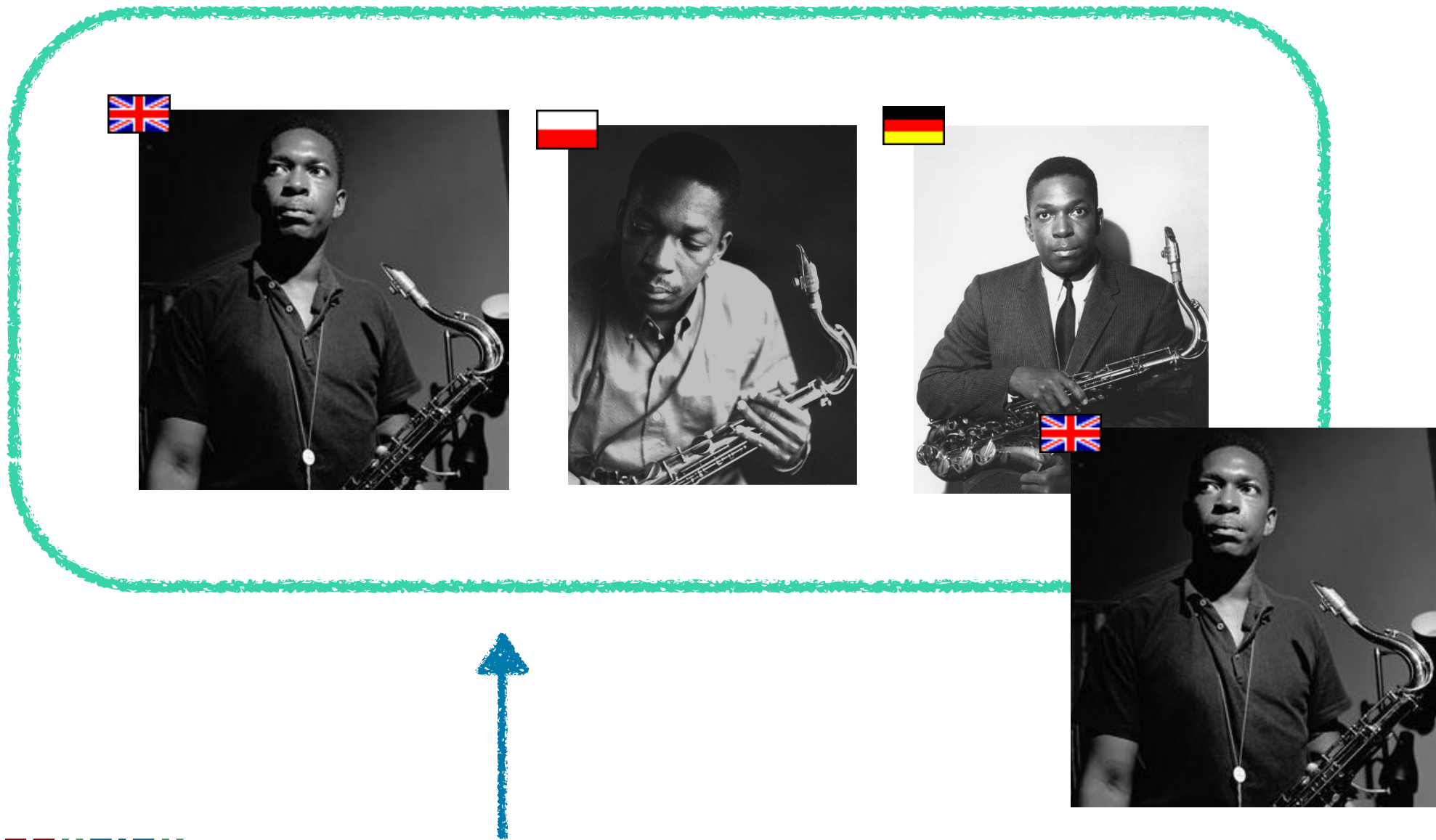


I. Set of Entities



specifies the abstraction classes

I. Set of Entities



specifies the abstraction classes

2. Taxonomy construction



en/John_Coltrane inCategory "Jazz Music"

en/John_Coltrane inCategory "American Composers"

2. Taxonomy construction

 en/John_Coltrane inCategory "Jazz Music"
en/John_Coltrane inCategory "American Composers"



 en/John_Coltrane type American_Composer

2. Taxonomy construction

 en/John_Coltrane inCategory "Jazz Music"
en/John_Coltrane inCategory "American Composers"

 en/John_Coltrane type American **Composer**

 American_Composer subclassOf wordnet **composer**

2. Taxonomy construction

English-centric!

 en/John_Coltrane inCategory "Jazz Music"
en/John_Coltrane inCategory "American Composers"

 en/John_Coltrane type American_Composer

 American_Composer subclassOf wordnet_composer

2. Taxonomy construction



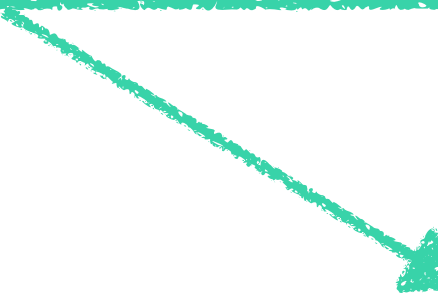
pl/John_Coltrane inCategory pl/Amerykańscy_Jazzmani



en/John_Coltrane inCategory "Jazz Music"
en/John_Coltrane inCategory "American Composers"



en/John_Coltrane type American_Composer



American_Composer subclassOf wordnet_composer

2. Taxonomy construction

 pl/John_Coltrane inCategory pl/Amerykańscy_Jazzmani



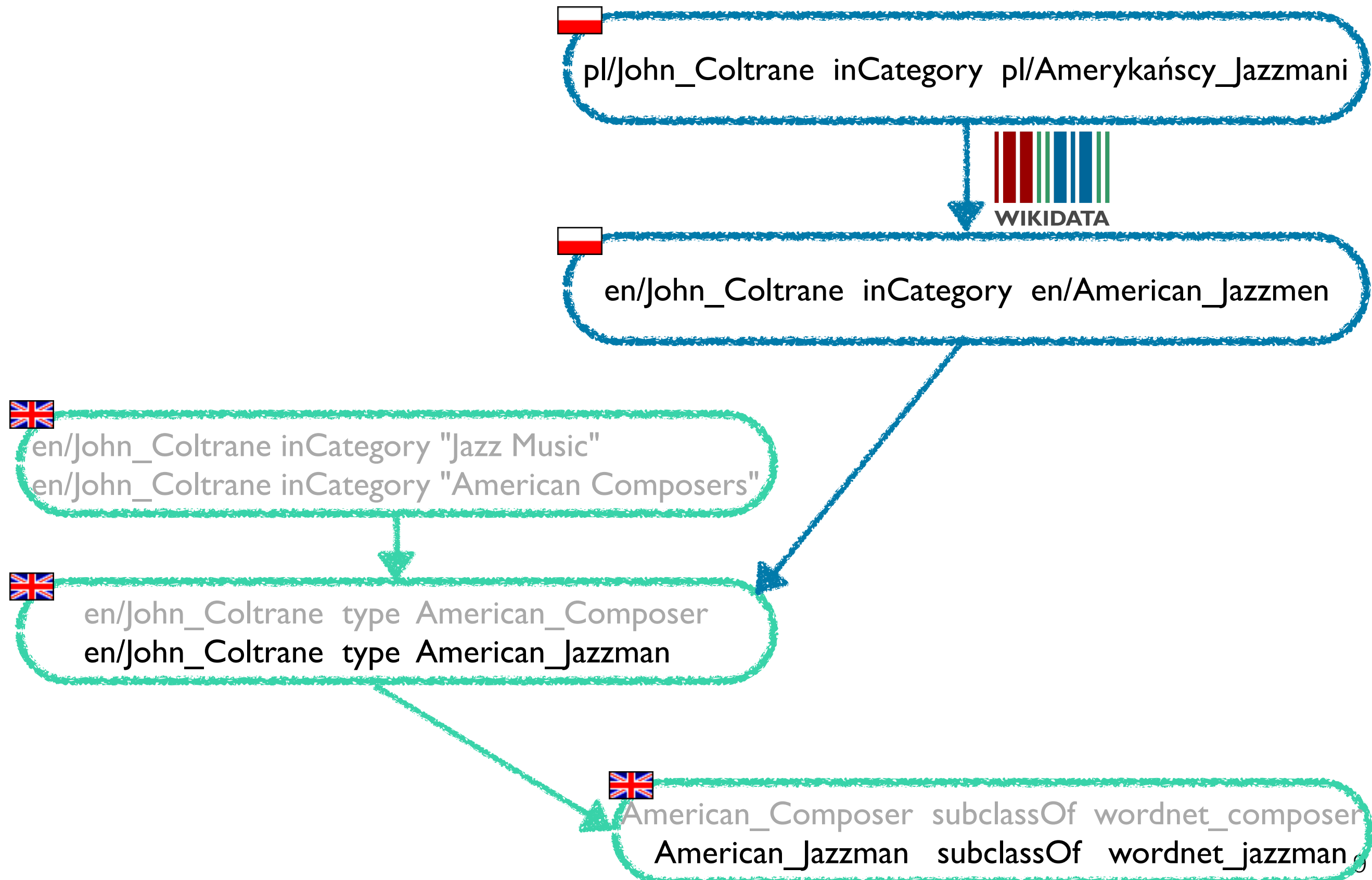
 en/John_Coltrane inCategory en/American_Jazzmen

 en/John_Coltrane inCategory "Jazz Music"
en/John_Coltrane inCategory "American Composers"

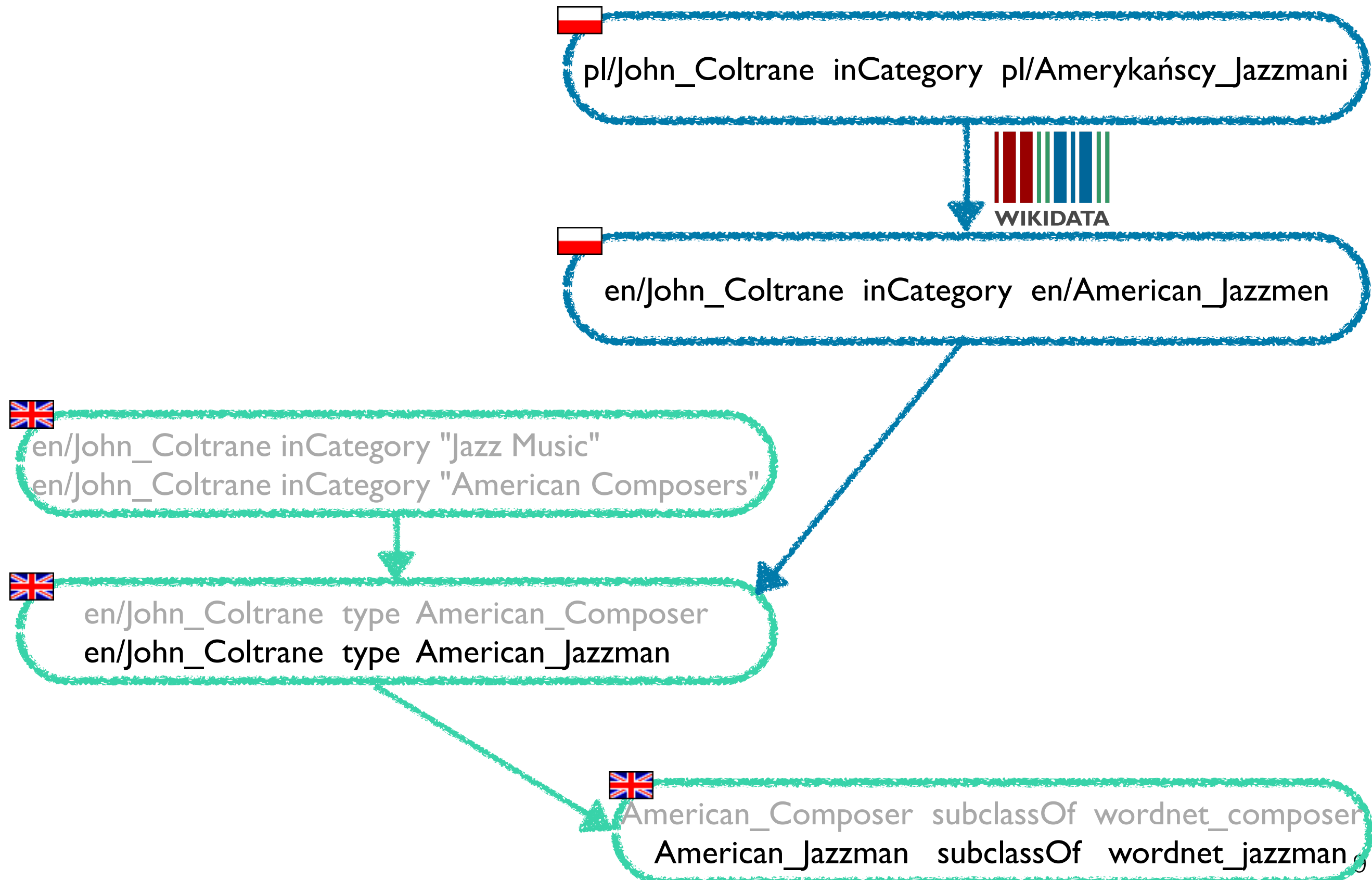
 en/John_Coltrane type American_Composer

 American_Composer subclassOf wordnet_composer

2. Taxonomy construction



2. Taxonomy construction



3. Fact extraction



en/infobox/married



3. Fact extraction



en/infobox/married



isMarriedTo



Manually defined in YAGO-EN

3. Fact extraction



en/infobox/married



isMarriedTo



pl/infobox/małżonek



3. Fact extraction



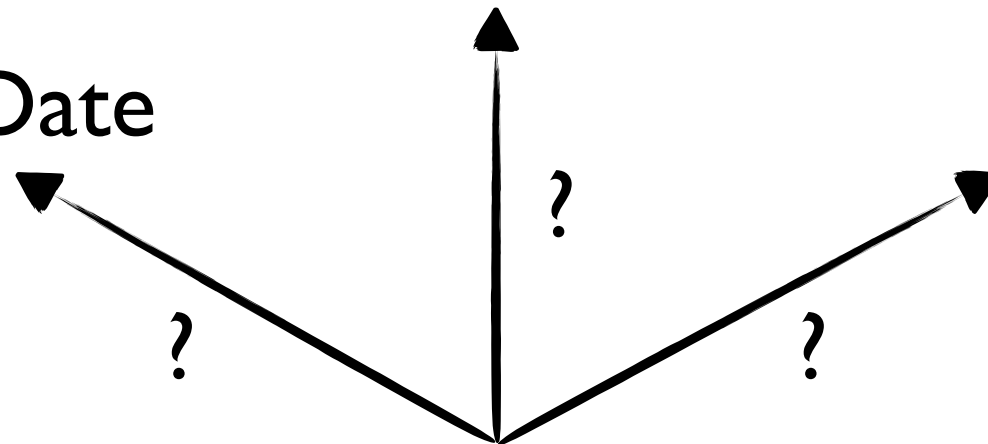
en/infobox/married



isMarriedTo

wasBornOnDate

hasChild



pl/infobox/małżonek



Infobox attributes mapping

pl/infobox/małżonek

=?

isMarriedTo

$F_{malzonek}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{małzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

Corresponding attributes will share
some subject-object pairs

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)

(Elvis_Presley, Priscilla_Presley)

(John_Coltrane, Ravi_Coltrane)

(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

(Barack_Obama, Michelle_Obama)

(Elvis_Presley, Priscilla_Presley)

(John_Coltrane, Alice_Coltrane)

$$\text{support}(F_a, E_r) = |\text{matches}(F_a, E_r)|$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

Too restrictive for attributes with few contributions

$$\text{support}(F_a, E_r) = |\text{matches}(F_a, E_r)|$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

$$confidence(F_a, E_r) = \frac{|matches(F_a, E_r)|}{|contrib(F_a)|}$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
 (Elvis_Presley, Priscilla_Presley)
 (John_Coltrane, Ravi_Coltrane)
 (pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)
 pl/Krystyna_Pyrkosz, pl/Witold_Pyrkosz
 pl/Grażyna_Torbicka, pl/Adam_Torbicki
 pl/Szymon_Majewski, pl/Magda_Majewska

(Barack_Obama, Michelle_Obama)
 (Elvis_Presley, Priscilla_Presley)
 (John_Coltrane, Alice_Coltrane)

Too restrictive for attributes
 with a lot of new facts
 but few matches

$$confidence(F_a, E_r) = \frac{|matches(F_a, E_r)|}{|contrib(F_a)|}$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)
pl/Krystyna_Pyrkosz, pl/Witold_Pyrkosz
pl/Grażyna_Torbicka, pl/Adam_Torbicki
pl/Szymon_Majewski, pl/Magda_Majewska

Open-world assumption

$$pca(F_a, E_r) = \frac{|matches(F_a, E_r)|}{|matches(F_a, E_r)| + |clashes(F_a, E_r)|}$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
 (Elvis_Presley, Priscilla_Presley)
 (John_Coltrane, Ravi_Coltrane)
 (pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)
 pl/Krystyna_Pyrkosz, pl/Witold_Pyrkosz
 pl/Grażyna_Torbicka, pl/Adam_Torbicki
 pl/Szymon_Majewski, pl/Magda_Majewska

(Barack_Obama, Michelle_Obama)
 (Elvis_Presley, Priscilla_Presley)
 (John_Coltrane, Alice_Coltrane)

Can get mislead by clashes

Open-world assumption

$$pca(F_a, E_r) = \frac{|matches(F_a, E_r)|}{|matches(F_a, E_r)| + |clashes(F_a, E_r)|}$$

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

...

$F_{malzonek}^*$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

Random Sample

Infobox attributes mapping

pl/infobox/małzonek

=?

isMarriedTo

$F_{malzonek}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Ravi_Coltrane)
(pl/Izabella_Olszewska, pl/Tadeusz_Jurasz)

...

$F_{malzonek}^*$

$E_{isMarriedTo}$

(Barack_Obama, Michelle_Obama)
(Elvis_Presley, Priscilla_Presley)
(John_Coltrane, Alice_Coltrane)

With 95% probability
the true proportion
of matches falls into

$[c - \delta, c + \delta]$

$wilson(F_a, E_r) = c - \delta$

3. Fact extraction



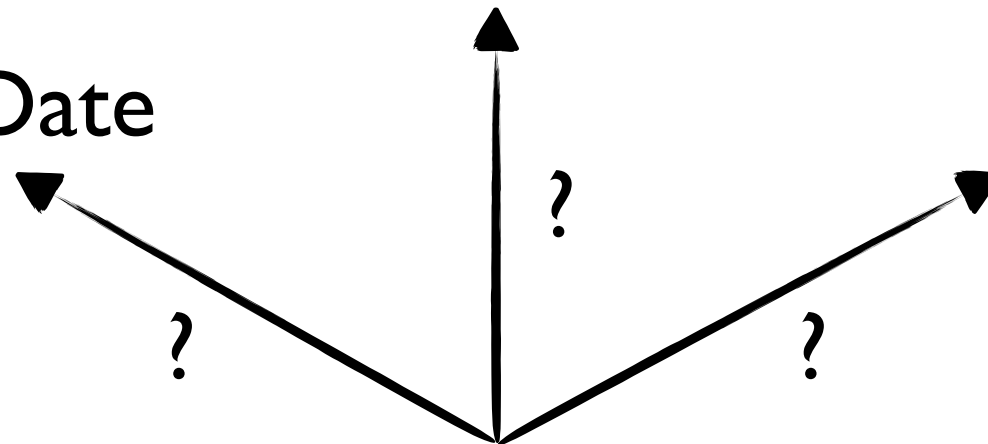
en/infobox/married



isMarriedTo

wasBornOnDate

hasChild



pl/infobox/małżonek



3. Fact extraction



en/infobox/married



isMarriedTo



pl/infobox/małżonek



Mapping quality



Confidence 16%	Wilson 4%
----------------	-----------

Estimated on a manually
annotated sample

Mapping quality

Confidence 16%

Wilson 4%



Good performance across
different languages

Mapping quality

Confidence 16%

Wilson 4%



Chosen so that we get high
recall at precision > 95%

Mapping quality

	Confidence 16%			Wilson 4%		
	Prec	Rec	F1	Prec	Rec	F1
ar	100	73	85	100	82	90
de	100	37	54	98	56	72
es	96	19	32	95	29	45
fa	100	49	66	97	54	69
fr	100	16	27	100	69	82
it	100	7	12	98	23	37
nl	100	19	32	100	22	36
pl	95	10	19	97	64	77
ro	96	52	67	95	70	81

Mapping quality

	Confidence 16%			Wilson 4%		
	Prec	Rec	F1	Prec	Rec	F1
ar	100	73	85	100	82	90
de	100	37	54	98	56	72
es	96	19	32	95	29	45
fa	100	49	66	97	54	69
fr	100	16	27	100	69	82
it	100	7	12	98	23	37
nl	100	19	32	100	22	36
pl	95	10	19	97	64	77
ro	96	52	67	95	70	81

High precision consistent across languages.

Mapping quality

	Confidence 16%			Wilson 4%		
	Prec	Rec	F1	Prec	Rec	F1
ar	100	73	85	100	82	90
de	100	37	54	98	56	72
es	96	19	32	95	29	45
fa	100	49	66	97	54	69
fr	100	16	27	100	69	82
it	100	7	12	98	23	37
nl	100	19	32	100	22	36
pl	95	10	19	97	64	77
ro	96	52	67	95	70	81

Higher recall for smaller wikipedias.

Mapping quality

	Confidence 16%				Wilson 4%		
	Prec	Rec	F1		Prec	Rec	F1
ar	100	73	85	<	100	82	90
de	100	37	54		98	56	72
es	96	19	32		95	29	45
fa	100	49	66		97	54	69
fr	100	16	27		100	69	82
it	100	7	12		98	23	37
nl	100	19	32		100	22	36
pl	95	10	19		97	64	77
ro	96	52	67		95	70	81

Lower threshold for Wilson helps increase recall.

YAGO3



YAGO3



```
de/Kirdorf_(Bedburg),  
  hasNumberOfPeople, "1204"^^xsd:integer  
fr/Château_de_Montcony,  
  isLocatedIn, Burgundy  
pl/Henryk_Pietras, wasBornIn  
de/Debiensko
```

1M new entities (3.5M for English)

2.5M new facts (6.5M for English)

YAGO3



```
de/Kirdorf_(Bedburg),  
  hasNumberOfPeople, "1204"^^xsd:integer  
fr/Château_de_Montcony,  
  isLocatedIn, Burgundy  
pl/Henryk_Pietras, wasBornIn  
de/Debiensko
```

1M new entities (3.5M for English)

2.5M new facts (6.5M for English)

Large, clean knowledge base from multilingual wikipeidias.

Single coherent taxonomy.

Mapping of infobox attributes to YAGO relations.

YAGO3



Download

<http://yago-knowledge.org>

Thank you!