(Open) Information Extraction: Where are we going?

SAPIENZA
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory

Claudio Delli Bovi
July 18th, 2016

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

About me

dellibovi@di.uniroma1.it ✉

`http://wwwusers.di.uniroma1.it/~dellibovi` 🏠

bn:17381128n

Second-year PhD student

LCL group @ Sapienza

Advisor: prof. Roberto Navigli

Focus (so far): Disambiguation, (Open) Information Extraction

Outline

BabelNet and friends: some background
Research work @ LCL Sapienza

DefIE: OIE from textual definitions
Delli Bovi, Telesca, Navigli: **TACL 2015**

KBUnify: KB disambiguation and unification
Delli Bovi, Espinosa-Anke, Navigli: **EMNLP 2015**

Outline

BabelNet and friends: some background
Research work @ LCL Sapienza

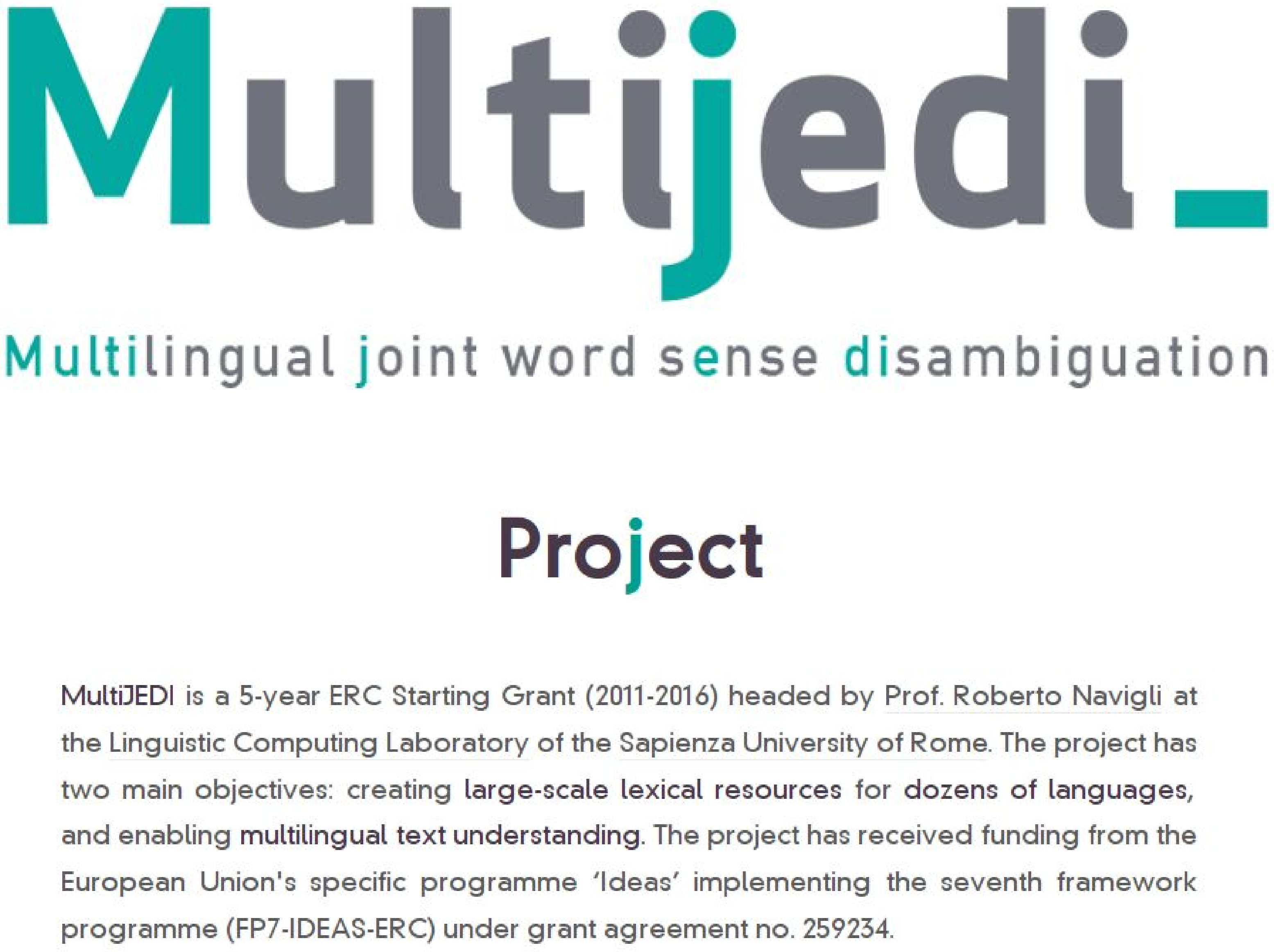
DefIE: OIE from textual definitions
Delli Bovi, Telesca, Navigli: **TACL 2015**

KBUnify: KB disambiguation and unification
Delli Bovi, Espinosa-Anke, Navigli: **EMNLP 2015**

Linguistic Computing Laboratory (LCL) @ Sapienza University of Rome

- Part of the **Computer Science Department** of Sapienza, focused on **Natural Language Processing**
- Some projects we have been involved in:
  - **MultiJEDI (1.3M €)**: ERC Starting Grant
  - **LIDER (1.5 M €)**: EU CSA
  - **Google Focused Research Award (300k $)**

http://multijedi.org/



Multijedi_

Multilingual joint word sense disambiguation

Project

MultiJEDI is a 5-year ERC Starting Grant (2011-2016) headed by Prof. Roberto Navigli at the Linguistic Computing Laboratory of the Sapienza University of Rome. The project has two main objectives: creating large-scale lexical resources for dozens of languages, and enabling multilingual text understanding. The project has received funding from the European Union's specific programme 'Ideas' implementing the seventh framework programme (FP7-IDEAS-ERC) under grant agreement no. 259234.

BabelNet

BabelNet

- To the best of our knowledge, the largest **multilingual encyclopedic dictionary** and **semantic network** (almost **14M** entries in **271** languages and **380M** semantic connections)

BabelNet

- To the best of our knowledge, the largest **multilingual encyclopedic dictionary** and **semantic network** (almost **14M** entries in **271** languages and **380M** semantic connections)

- Initially created as an integration of **Wikipedia** and **WordNet**, now BabelNet is a merger of many different resources (Wiktionary, Wikidata, OmegaWiki, VerbNet, ImageNet, ...)

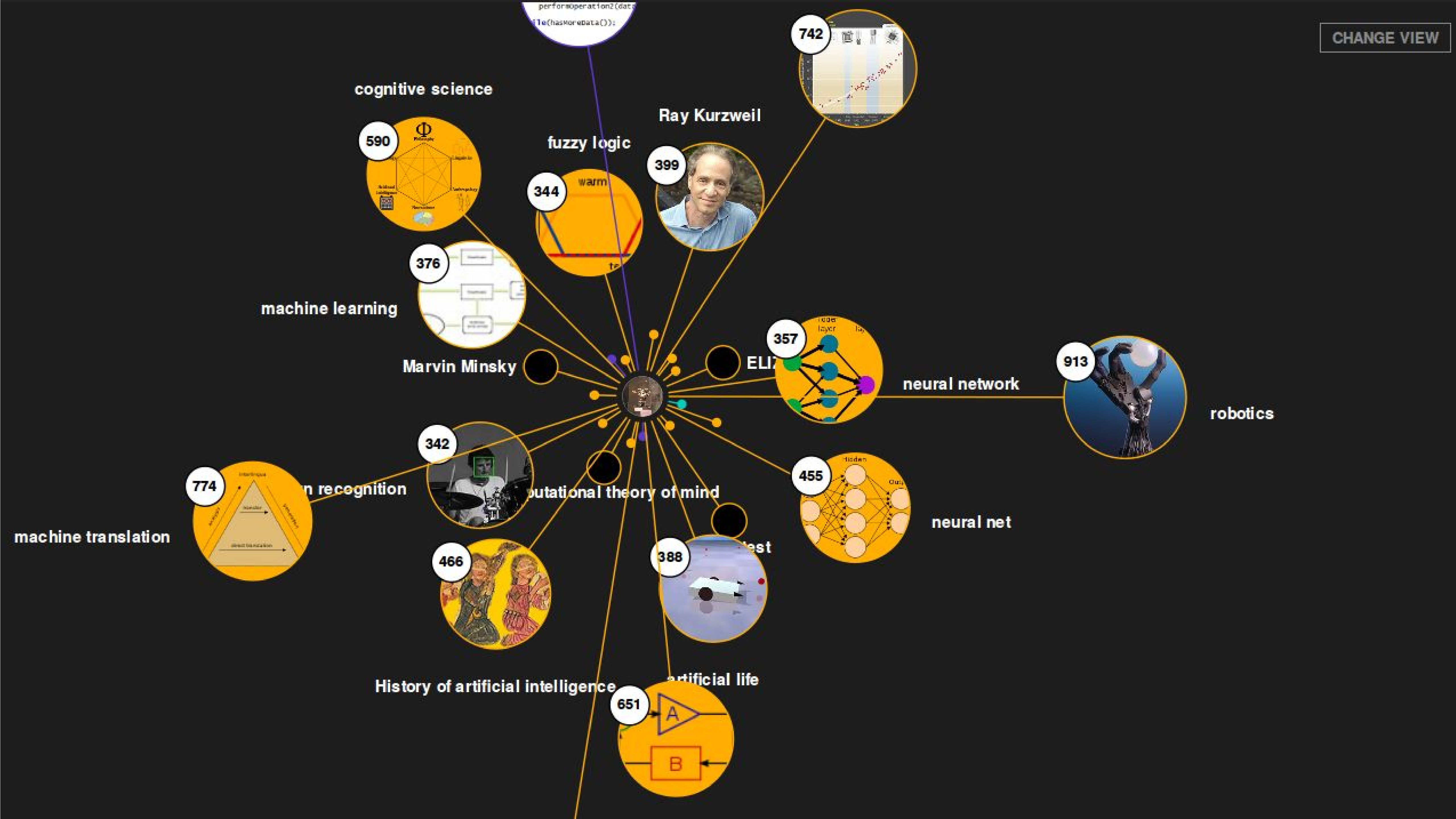
BabelNet

- The integration is performed via an **automatic linking algorithm** and by filling in lexical gaps with the aid of **Machine Translation**

BabelNet

- The integration is performed via an **automatic linking algorithm** and by filling in lexical gaps with the aid of **Machine Translation**

- BabelNet is composed of **Babel Synsets**, concepts or entities **lexicalized** ("WordNet-style") in many languages and featuring:
  - **is-a** relations
  - **domain** and **categories**
  - **images** and **definitions**
  - **translations**


performOperation2(data
ile(hasMoreData());
CHANGE VIEW
742
cognitive science
Ray Kurzweil
fuzzy logic
590
399
344
warm
376
machine learning
357
Marvin Minsky
ELIZ
neural network
913
robotics
342
774
n recognition
putational theory of mind
455
neural net
machine translation
466
388
test
History of artificial intelligence
artificial life
651
A
B

BabelNet and friends

BabelNet and friends

Babelfy

A graph-based algorithm for multilingual joint **Word Sense Disambiguation** and **Entity Linking**, based on BabelNet

BabelNet and friends

Babelfy

A graph-based algorithm for multilingual joint **Word Sense Disambiguation** and **Entity Linking**, based on BabelNet

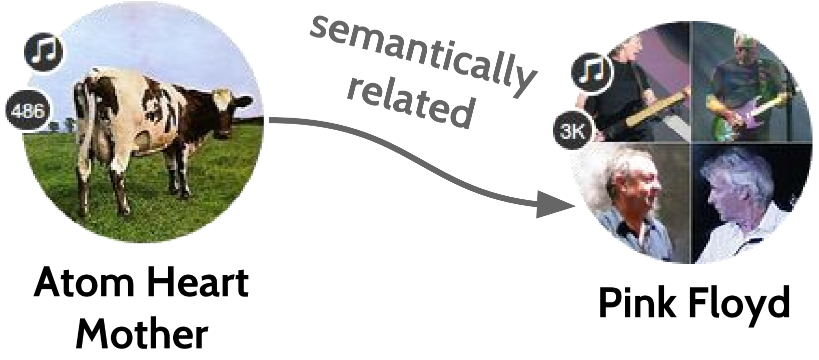
The Wikipedia Bitaxonomy

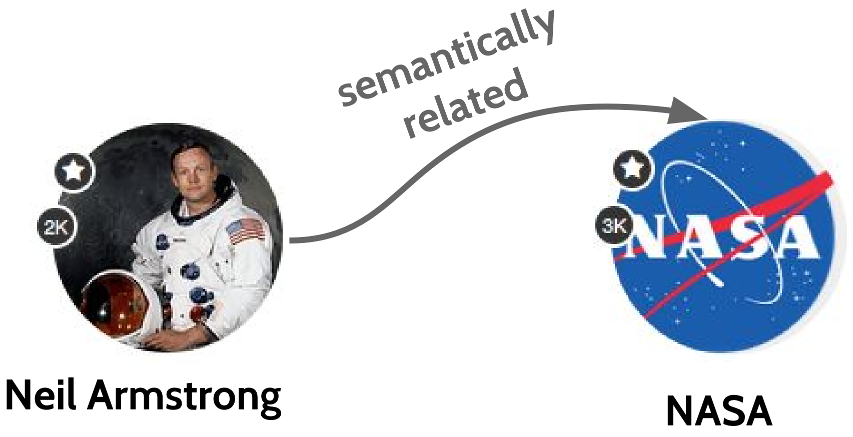
An iterative algorithm for the automatic creation of a “**bitaxonomy**” for Wikipedia pages and categories

... and much more!

BabelNet and my research

- BabelNet (especially in its early stages) was conceived as a **lexico-semantic resource** more than an actual **knowledge base**:
  - semantic connections are mostly **lexical relations** from WordNet or unspecified "**relatedness edges**" derived from Wikipedia hyperlinks



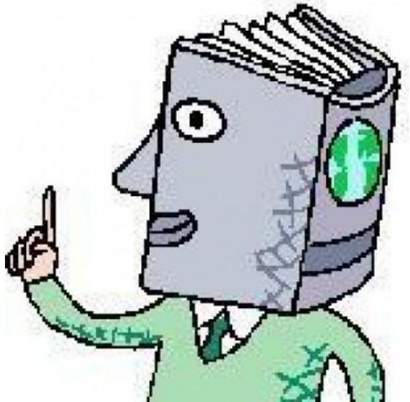


BabelNet and my research

- BabelNet (especially in its early stages) was conceived as a **lexico-semantic resource** more than an actual **knowledge base**:
  - semantic connections are mostly **lexical relations** from WordNet or unspecified “**relatedness edges**” derived from Wikipedia hyperlinks

- Construct from BabelNet a proper knowledge base with **labeled relations** (X is album by Y, X worked at Y, ...)

- Use **Open Information Extraction**!

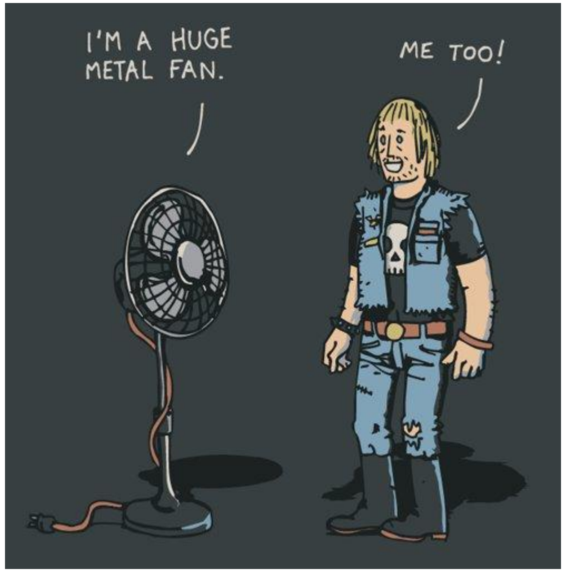


(Open) Information Extraction

OIE is great, but...

Sparsity: many relation phrases express the same relationship (e.g. synonyms, paraphrases)

Ambiguity: arguments (and relation phrases) are ambiguous!



Outline

BabelNet and friends: some background
Research work @ LCL Sapienza

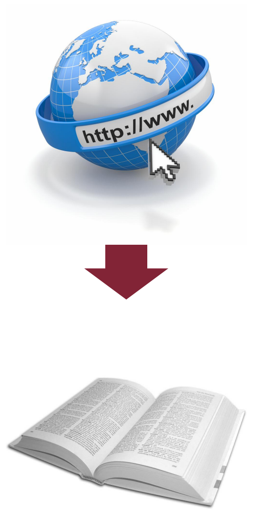
DefIE: OIE from textual definitions
Delli Bovi, Telesca, Navigli: **TACL 2015**

KBUnify: KB disambiguation and unification
Delli Bovi, Espinosa-Anke, Navigli: **EMNLP 2015**

DefIE: OIE from textual definitions

The idea:

instead of targeting massive and noisy corpora (like the web) and then trying to find a smart way to cope with the noise

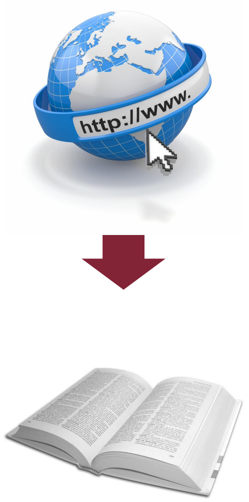


target smaller but “denser” (and virtually noise-free) corpora of **definitional knowledge**.

DefIE: OIE from textual definitions

The idea:

instead of targeting massive and noisy corpora (like the web) and then trying to find a smart way to cope with the noise



target smaller but “denser” (and virtually noise-free) corpora of **definitional knowledge**.

Apply OIE techniques to extract as much information as possible!

DefIE: OIE from textual definitions

The tools:

- An underlying **inventory/knowledge base** (to which arguments and relation patterns will be connected)

- A **WSD/EL system** (to disambiguate concepts and entity mentions across the input text)

- A **syntactic parser** (to construct meaningful relation patterns and avoid sparsity)

DefIE: OIE from textual definitions

The tools:

- An underlying **inventory/knowledge base** (to which arguments and relation patterns will be connected)

- A **WSD/EL system** (to disambiguate concepts and entity mentions across the input text)

- A **syntactic parser** (to construct meaningful relation patterns and avoid sparsity)

http://babelnet.org

BabelNet

14 million entries

both **lexicographic** and **encyclopedic** knowledge

DefIE: OIE from textual definitions

The tools:

- An underlying **inventory/knowledge base** (to which arguments and relation patterns will be connected)

- A **WSD/EL system** (to disambiguate concepts and entity mentions across the input text)

- A **syntactic parser** (to construct meaningful relation patterns and avoid sparsity)

http://babelfy.org

Babelfy

unified graph-based approach to **EL** and **WSD**

unsupervised, based on **BabelNet**

DefIE: OIE from textual definitions

The tools:

- An underlying **inventory/knowledge base** (to which arguments and relation patterns will be connected)

- A **WSD/EL system** (to disambiguate concepts and entity mentions across the input text)

- A **syntactic parser** (to construct meaningful relation patterns and avoid sparsity)

http://svn.ask.it.usyd.edu.au/trac/candc

C&C tools

log-linear parser and supertagger based on **CCG**

(theoretically) suited to **long-distance dependencies**

DefIE: How it works

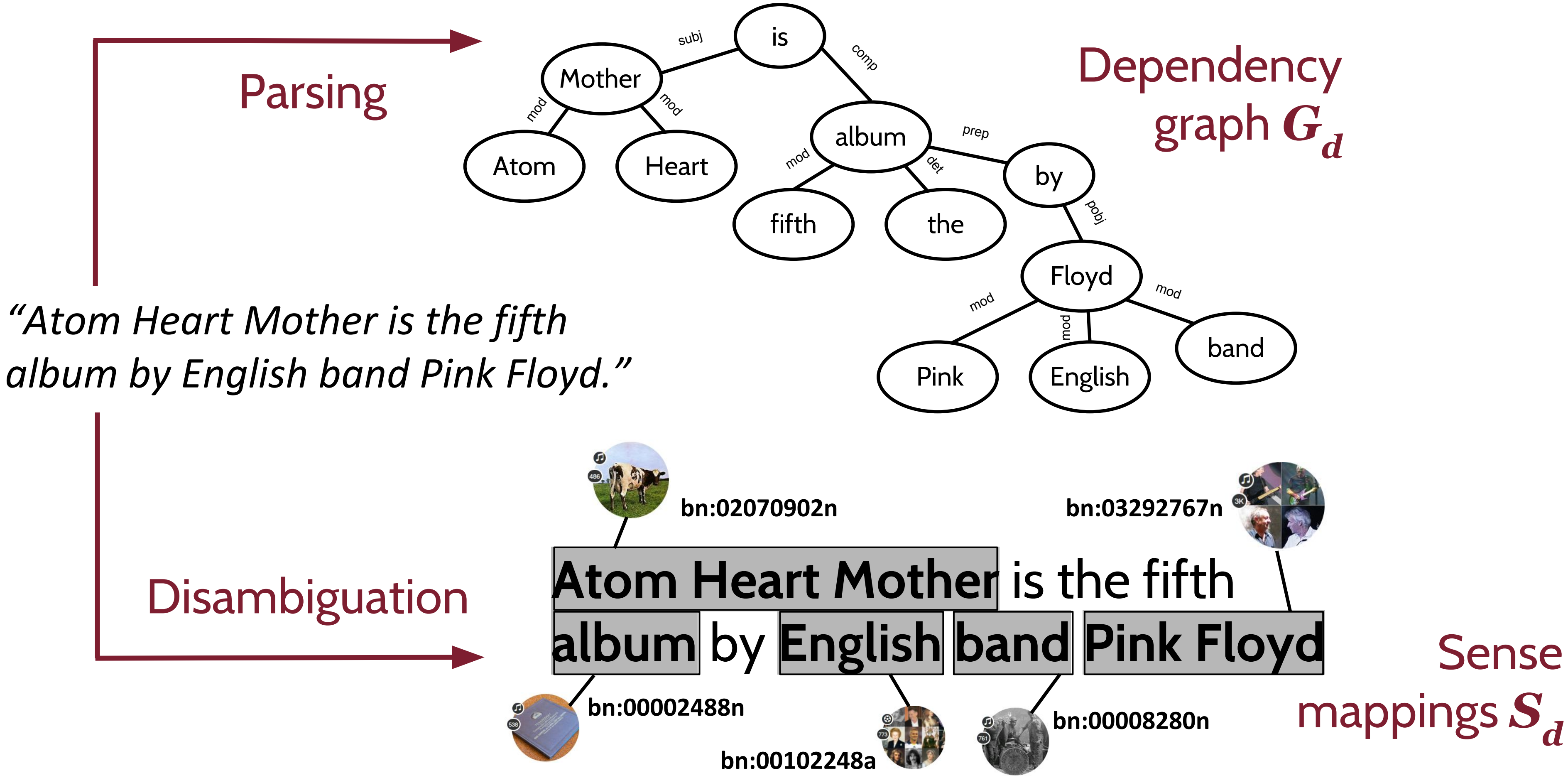
1. Extracting relation instances

“Atom Heart Mother is the fifth album by English band Pink Floyd.”

Textual definition $\boldsymbol{d}$

DefIE: How it works

1. Extracting relation instances



“Atom Heart Mother is the fifth album by English band Pink Floyd.”

Parsing

Dependency graph G_d

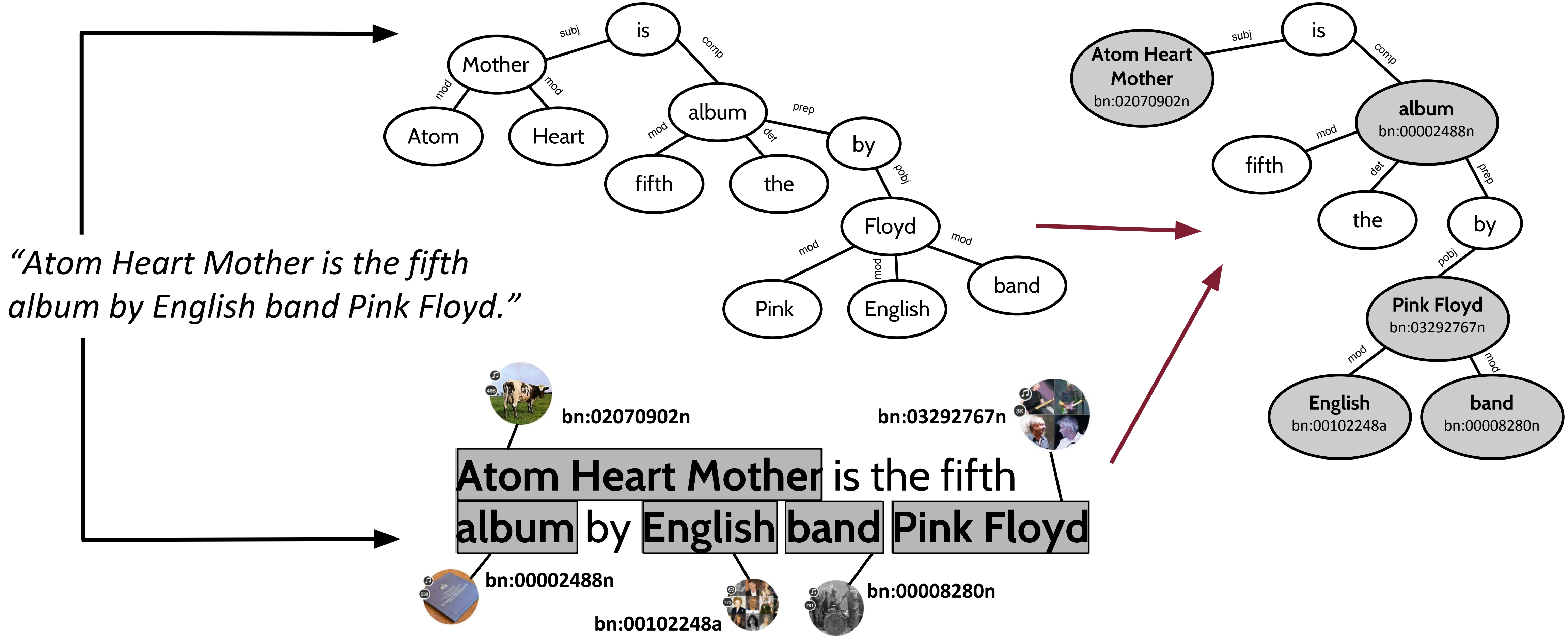
Disambiguation

Sense mappings S_d

DefIE: How it works

1. Extracting relation instances

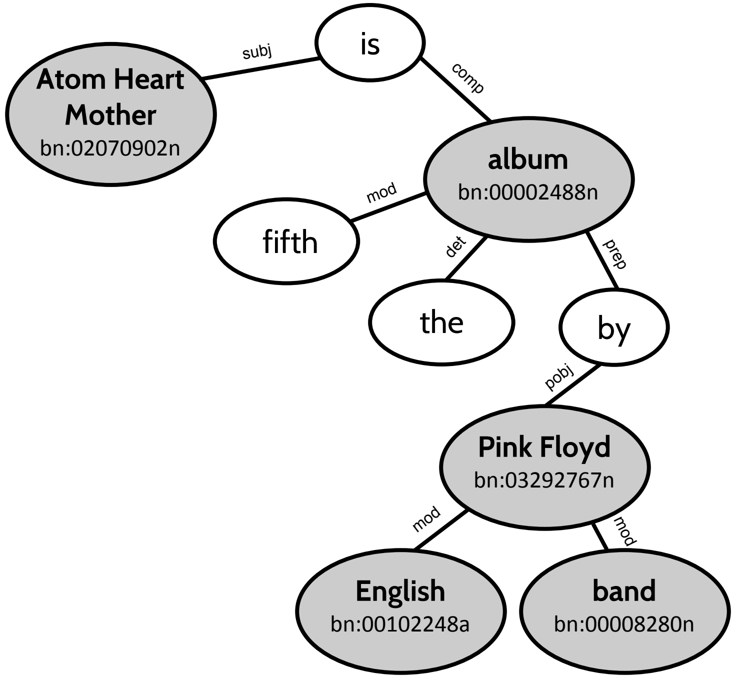
Syntactic-Semantic Graph S_d^{sem}



“Atom Heart Mother is the fifth album by English band Pink Floyd.”

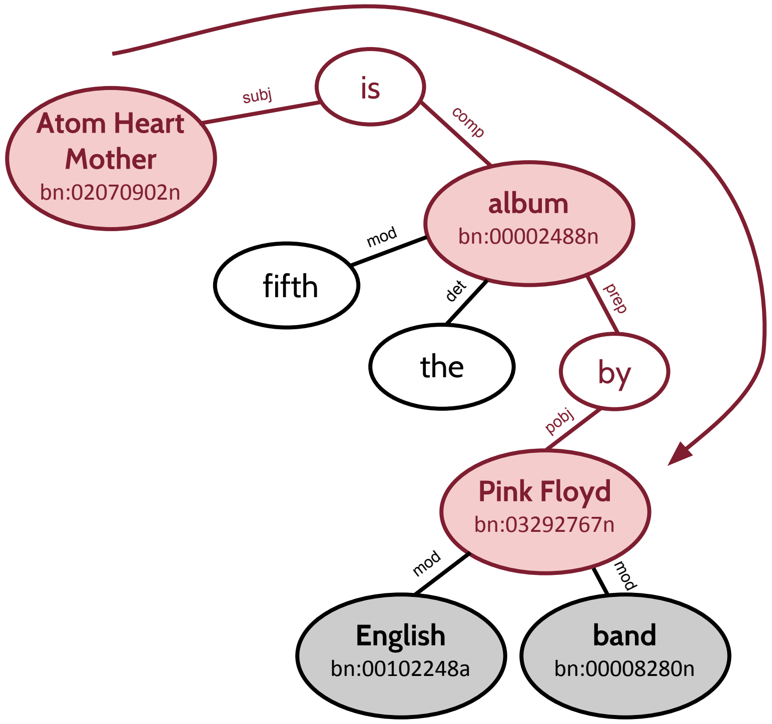
DefIE: How it works

1. Extracting relation instances



DefIE: How it works

1. Extracting relation instances



Extraction 1

$X \rightarrow \mathrm{is} \rightarrow$ **album** $\rightarrow \mathrm{by} \rightarrow Y$

bn:00002488n

X = **Atom Heart Mother** bn:02070902n

Y = **Pink Floyd** bn:03292767n

DefIE: How it works

1. Extracting relation instances

Extraction 2

$X \rightarrow \mathrm{is} \rightarrow Y$

X = **Atom Heart Mother** bn:02070902n

Y = **album** bn:00002488n

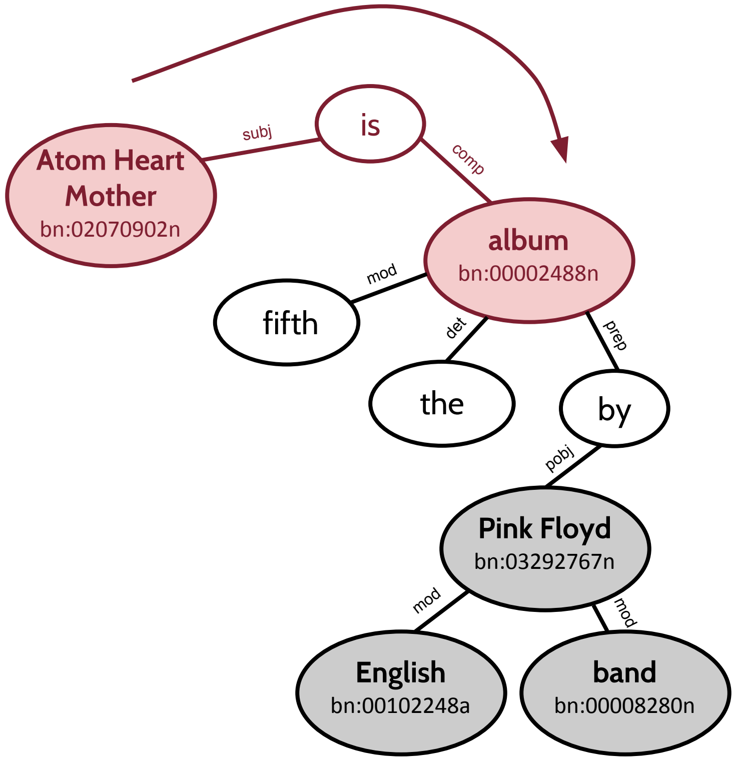
Extraction 1

$X \rightarrow \mathrm{is} \rightarrow$ **album** $\rightarrow \mathrm{by} \rightarrow Y$

bn:00002488n

X = **Atom Heart Mother** bn:02070902n

Y = **Pink Floyd** bn:03292767n



DefIE: How it works

1. Extracting relation instances

R_1: $\mathrm{X} \rightarrow \mathtt{is} \rightarrow \mathrm{Y}$ ➡

⟨Atom Heart Mother, album⟩
⟨Pink Floyd, band⟩
⋮
⟨Seattle, city⟩

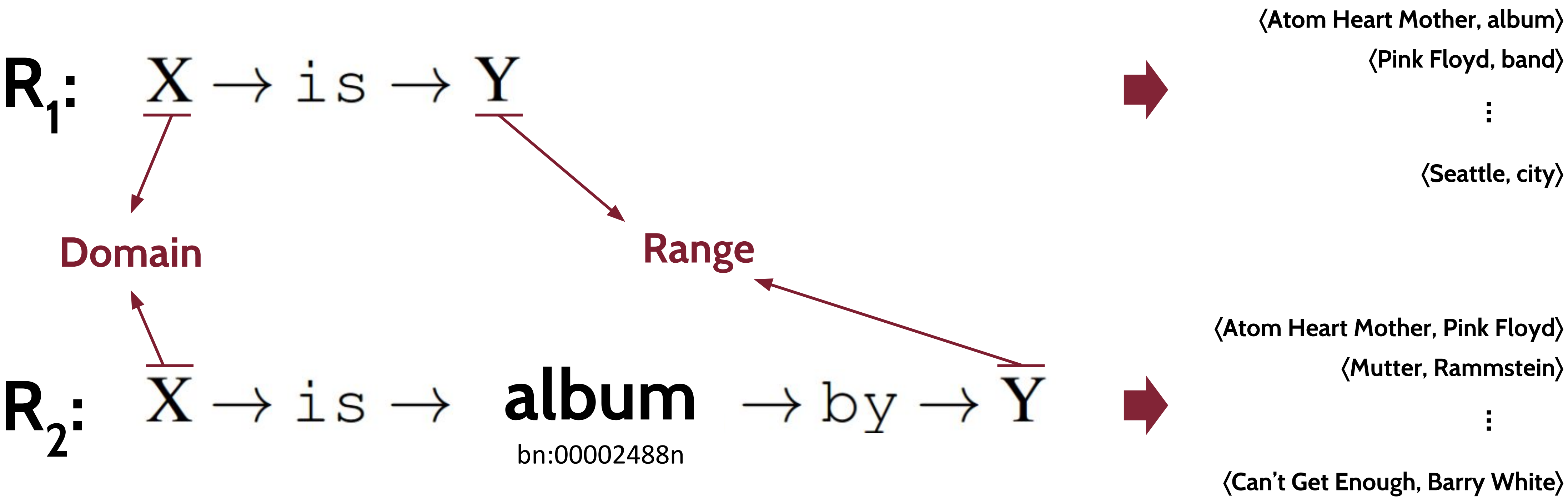
R_2: $\mathrm{X} \rightarrow \mathtt{is} \rightarrow$ **album** $\rightarrow \mathtt{by} \rightarrow \mathrm{Y}$ ➡

bn:00002488n

⟨Atom Heart Mother, Pink Floyd⟩
⟨Mutter, Rammstein⟩
⋮
⟨Can't Get Enough, Barry White⟩

DefIE: How it works

1. Extracting relation instances



R_1: $X \rightarrow \text{is} \rightarrow Y$

Domain

Range

R_2: $X \rightarrow \text{is} \rightarrow$ **album** $\rightarrow \text{by} \rightarrow Y$

bn:00002488n

⟨Atom Heart Mother, album⟩
⟨Pink Floyd, band⟩
⋮
⟨Seattle, city⟩

⟨Atom Heart Mother, Pink Floyd⟩
⟨Mutter, Rammstein⟩
⋮
⟨Can't Get Enough, Barry White⟩

DefIE: How it works

2. Relation typing and scoring

DefIE: How it works

2. Relation typing and scoring

For each relation R:

Substitute each domain and range argument with its **hypernym h** (using the BabelNet taxonomy) and generate a **probability distribution over semantic types** for the two sets

Compute the **entropy** of R as $H_R = -\sum_{i=1}^{n} p(h_i)\, log_2\, p(h_i)$

DefIE: How it works

2. Relation typing and scoring

For each relation R:

Compute the **score** of R as

Total number of extracted instances for R

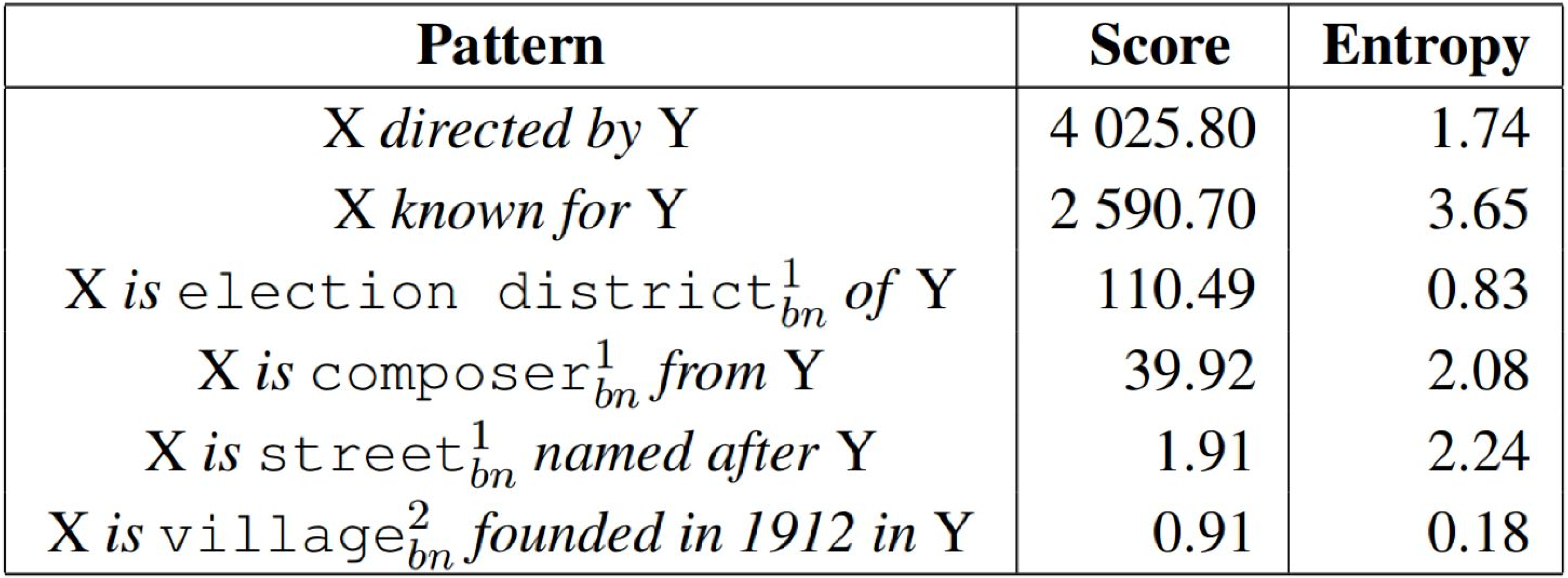
$$score(R) = \frac{|S_R|}{(H_R + 1)\, length(r)}$$

Domain and range entropy of R

Length of the relation pattern of R

DefIE: How it works

2. Relation typing and scoring



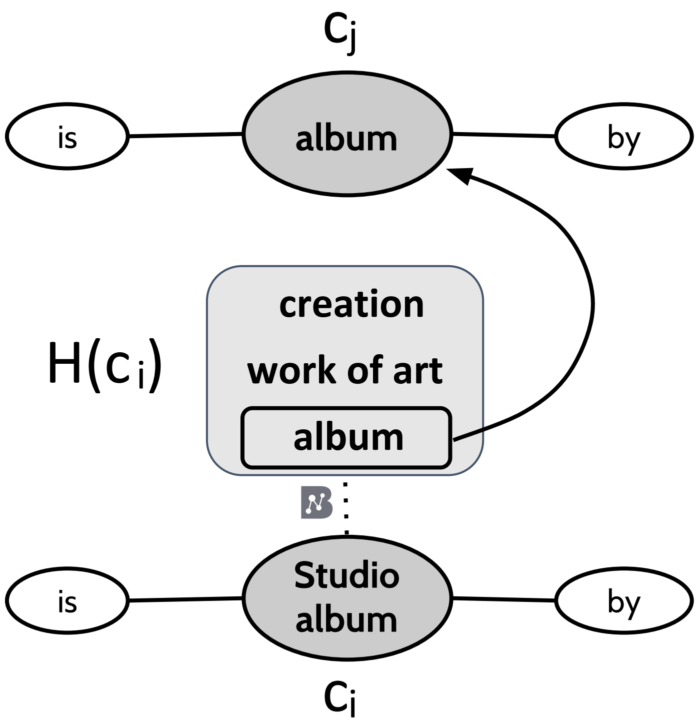
Pattern	Score	Entropy
X *directed by* Y	4 025.80	1.74
X *known for* Y	2 590.70	3.65
X *is* $\texttt{election district}_{bn}^{1}$ *of* Y	110.49	0.83
X *is* $\texttt{composer}_{bn}^{1}$ *from* Y	39.92	2.08
X *is* $\texttt{street}_{bn}^{1}$ *named after* Y	1.91	2.24
X *is* $\texttt{village}_{bn}^{2}$ *founded in 1912 in* Y	0.91	0.18

DefIE: How it works

3. Relation taxonomization

DefIE: How it works

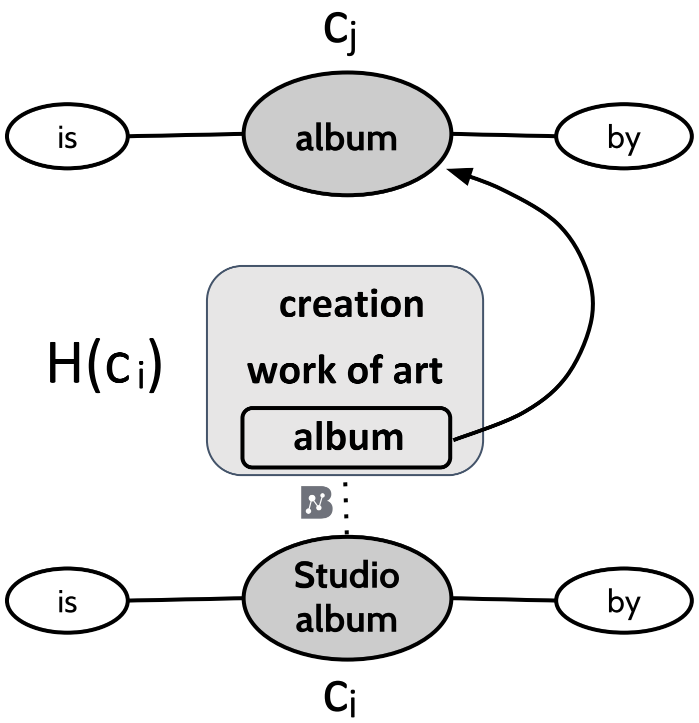
3. Relation taxonomization



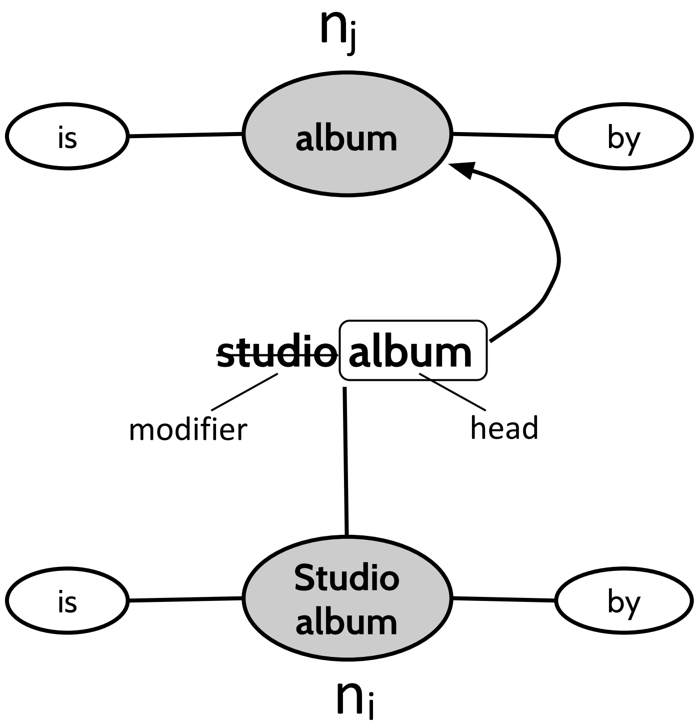
Hypernym generalization

DefIE: How it works

3. Relation taxonomization



Hypernym generalization



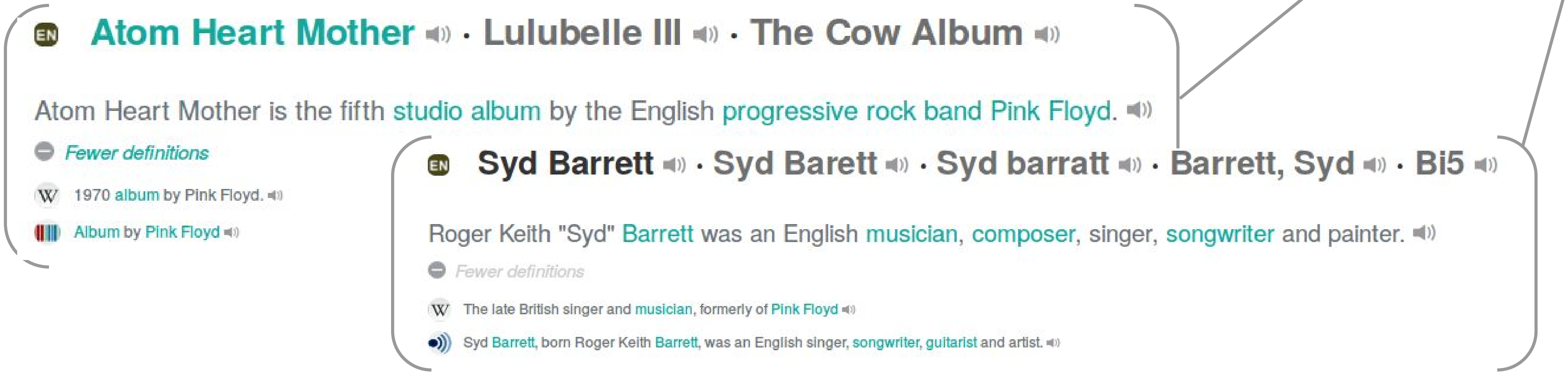
Substring generalization

DefIE: Setup

Dataset:
whole set of English textual definitions in BabelNet 2.5

4 357 327 items from **5** different sources (Wikipedia, WordNet, Wikidata, Wiktionary, OmegaWiki)

BabelNet



EN **Atom Heart Mother** · **Lulubelle III** · **The Cow Album**

Atom Heart Mother is the fifth studio album by the English progressive rock band Pink Floyd.

Fewer definitions

1970 album by Pink Floyd.

Album by Pink Floyd

EN **Syd Barrett** · **Syd Barett** · **Syd barratt** · **Barrett, Syd** · **Bi5**

Roger Keith "Syd" Barrett was an English musician, composer, singer, songwriter and painter.

Fewer definitions

The late British singer and musician, formerly of Pink Floyd

Syd Barrett, born Roger Keith Barrett, was an English singer, songwriter, guitarist and artist.

DefIE: Results

	DefIE	NELL	PATTY	ReVerb	WiSeNet
# Relations	255 881	298	**1 631 531**	664 746	245 935
Avg. extractions	81.68	**7 013.03**	9.68	22.16	9.24
# Extractions	**20 352 903**	2 089 883	15 802 946	14 728 268	2 271 807
# Entities	2 398 982	1 996 021	1 087 907	**3 327 425**	1 636 307
# Edges in the taxonomy	**44 412**	-	20 339	-	-

DefIE: Results

Other evaluations:

- **Precision** and **coverage** of relations
- **Novelty** of information
- Quality of relation **taxonomization**
- Quality of **entity linking/disambiguation**
- **Impact** of definition sources

...

DefIE: Future work

Where from here?

- Relation **clustering** (as in PATTY and WiSeNet)
- **Multilinguality**
- Relational **learning** and KB completion
- Harvest definitions from the **web**
- Adapt to “**general**” text

…

Outline

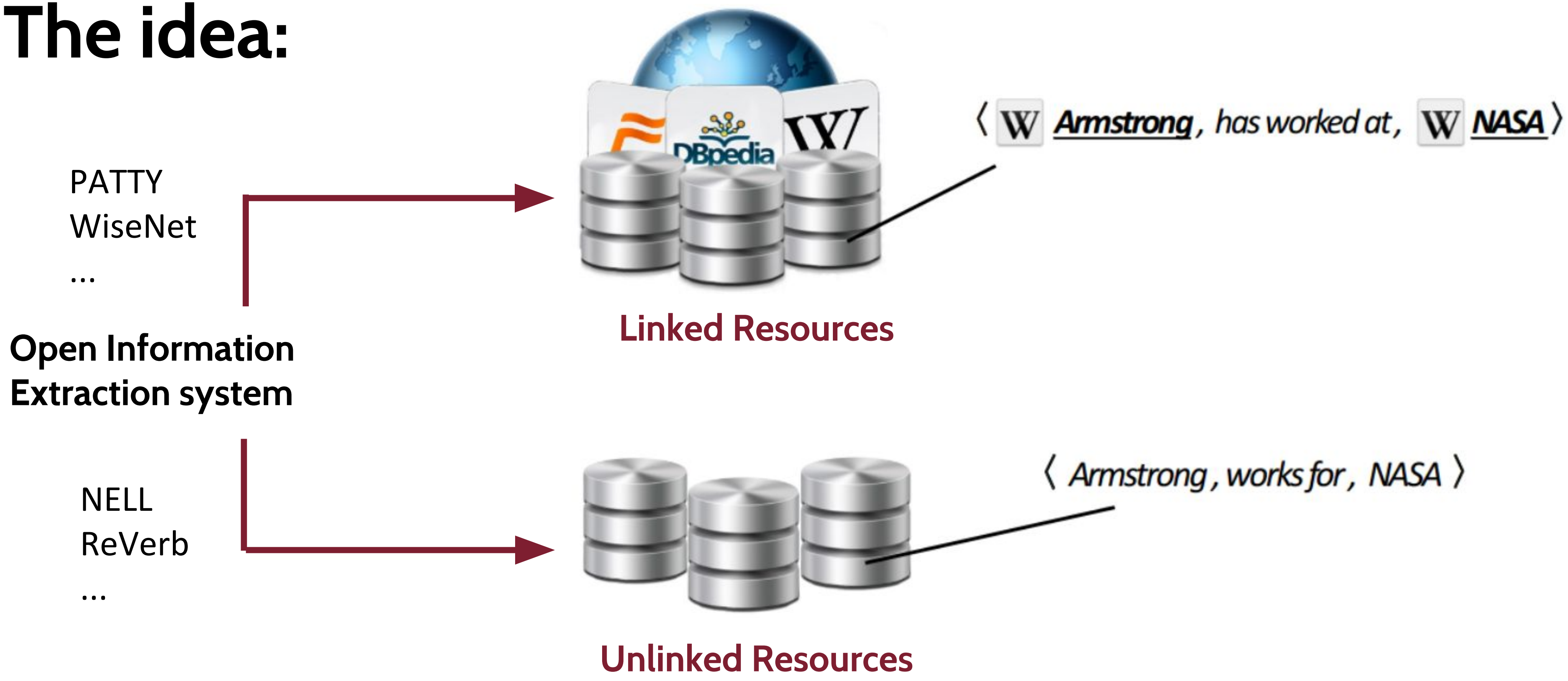
BabelNet and friends: some background
Research work @ LCL Sapienza

DefIE: OIE from textual definitions
Delli Bovi, Telesca, Navigli: **TACL 2015**

KBUnify: KB disambiguation and unification
Delli Bovi, Espinosa-Anke, Navigli: **EMNLP 2015**

KB-Unify: Knowledge base unification via sense embeddings and disambiguation

The idea:



PATTY
WiseNet
...

Open Information Extraction system

NELL
ReVerb
...

⟨ W *Armstrong*, *has worked at*, W *NASA* ⟩

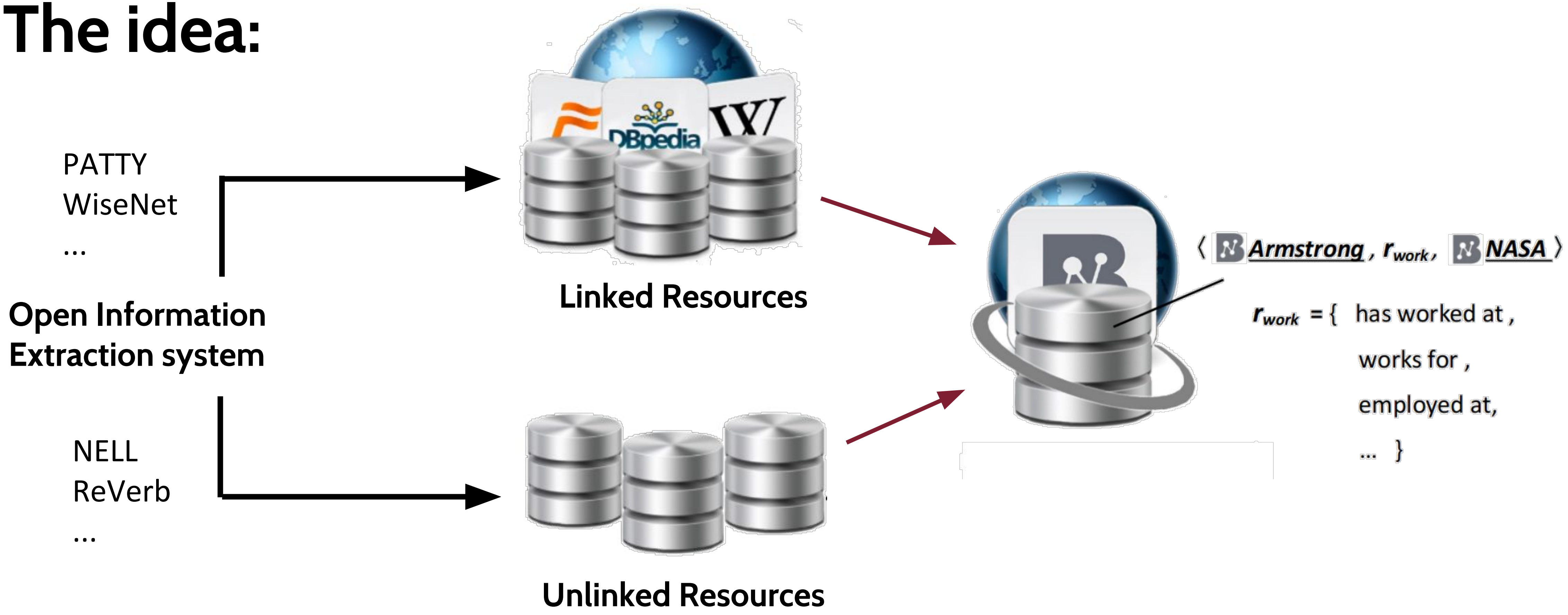
Linked Resources

⟨ *Armstrong*, *works for*, *NASA* ⟩

Unlinked Resources

KB-Unify: Knowledge base unification via sense embeddings and disambiguation

The idea:



PATTY
WiseNet
...

NELL
ReVerb
...

Open Information Extraction system

Linked Resources

Unlinked Resources

⟨ *Armstrong*, r_{work}, *NASA* ⟩

r_{work} = { has worked at ,
works for ,
employed at,
... }

KB-Unify: Knowledge base unification via sense embeddings and disambiguation

The tools:

- A **WSD/EL system** (to disambiguate unlinked resources)

- A unified **sense inventory S** (to make the various resources “speak to each other”)

- A unified **vector space V_S** (to associate a vector with each item of **S**)

KB-Unify: Knowledge base unification via sense embeddings and disambiguation

The tools:

- A **WSD/EL system** (to disambiguate unlinked resources)

- A unified **sense inventory S** (to make the various resources “speak to each other”)

- A unified **vector space V_S** (to associate a vector with each item of **S**)

Babelfy

Babelnet

KB-Unify: Knowledge base unification via sense embeddings and disambiguation

The tools:

- A **WSD/EL system** (to disambiguate unlinked resources)

- A unified **sense inventory S** (to make the various resources "speak to each other")

- A unified **vector space V_S** (to associate a vector with each item of **S**)

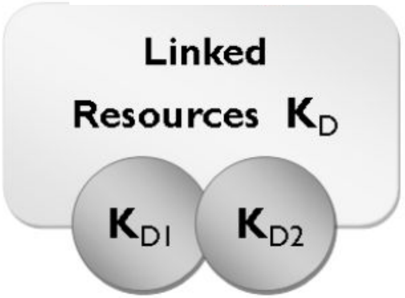
SensEmbed (Iacobacci et al., 2015)

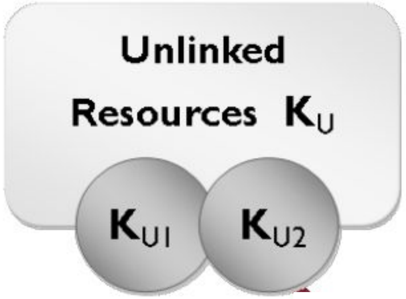
Sense-based embedding model

Popular word2vec architecture (**skip-gram**) trained on a **sense-annotated corpus**

KB-Unify: How it works

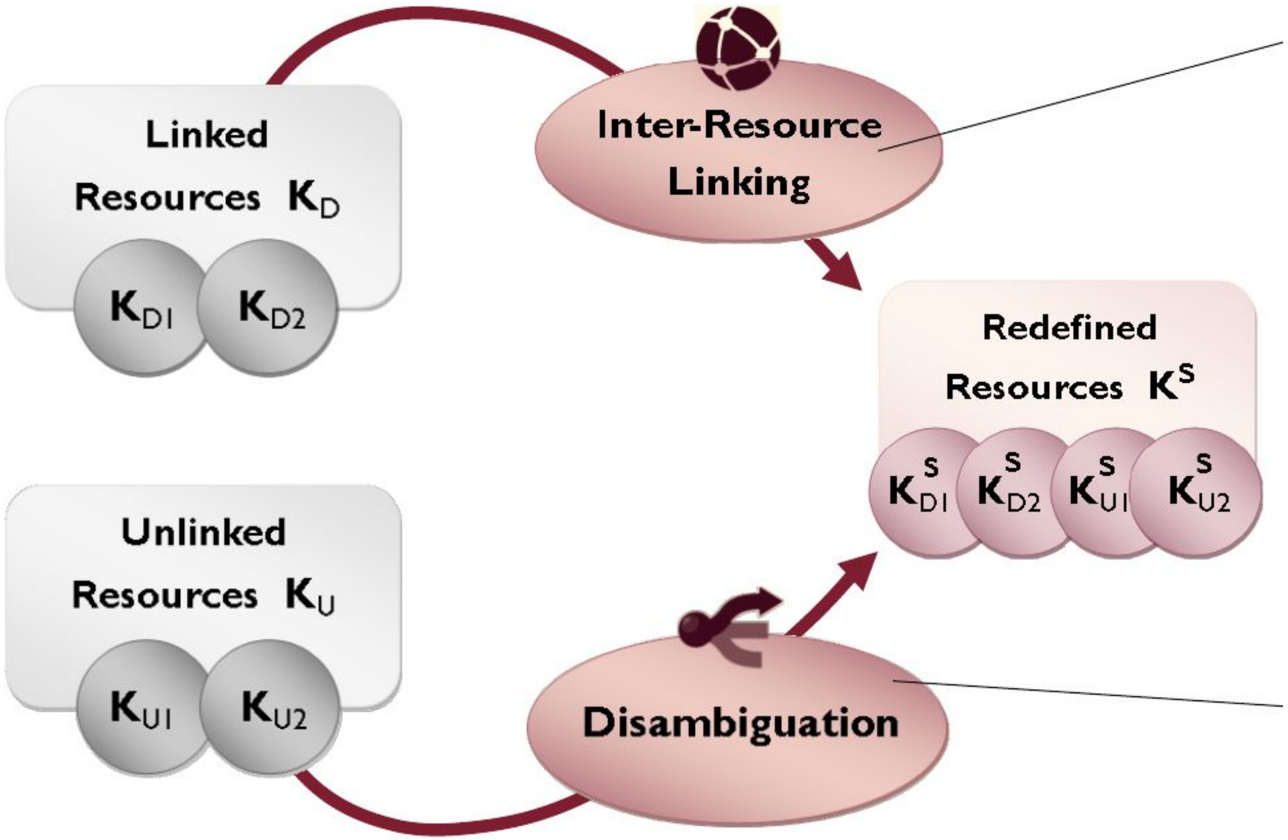
A bird's-eye view





KB-Unify: How it works

A bird's-eye view



use **BabelNet mappings** to redefine each linked resource

disambiguate each unlinked resource using BabelNet as sense inventory (more on this later!)

KB-Unify: How it works

- Disambiguation

KB-Unify: How it works

Disambiguation

Two basic intuitions:

1. Among all triples in target knowledge base, some of them (even if ambiguous) will be **easier to disambiguate**

   e.g. 〈 **Armstrong , works for , NASA** 〉

KB-Unify: How it works

Disambiguation

Two basic intuitions:

1. Among all triples in target knowledge base, some of them (even if ambiguous) will be **easier to disambiguate**

   e.g. ⟨ **Armstrong , works for , NASA** ⟩

2. In general, the disambiguation strategy should vary according to the **degree of specificity** of each relation

KB-Unify: How it works

Disambiguation

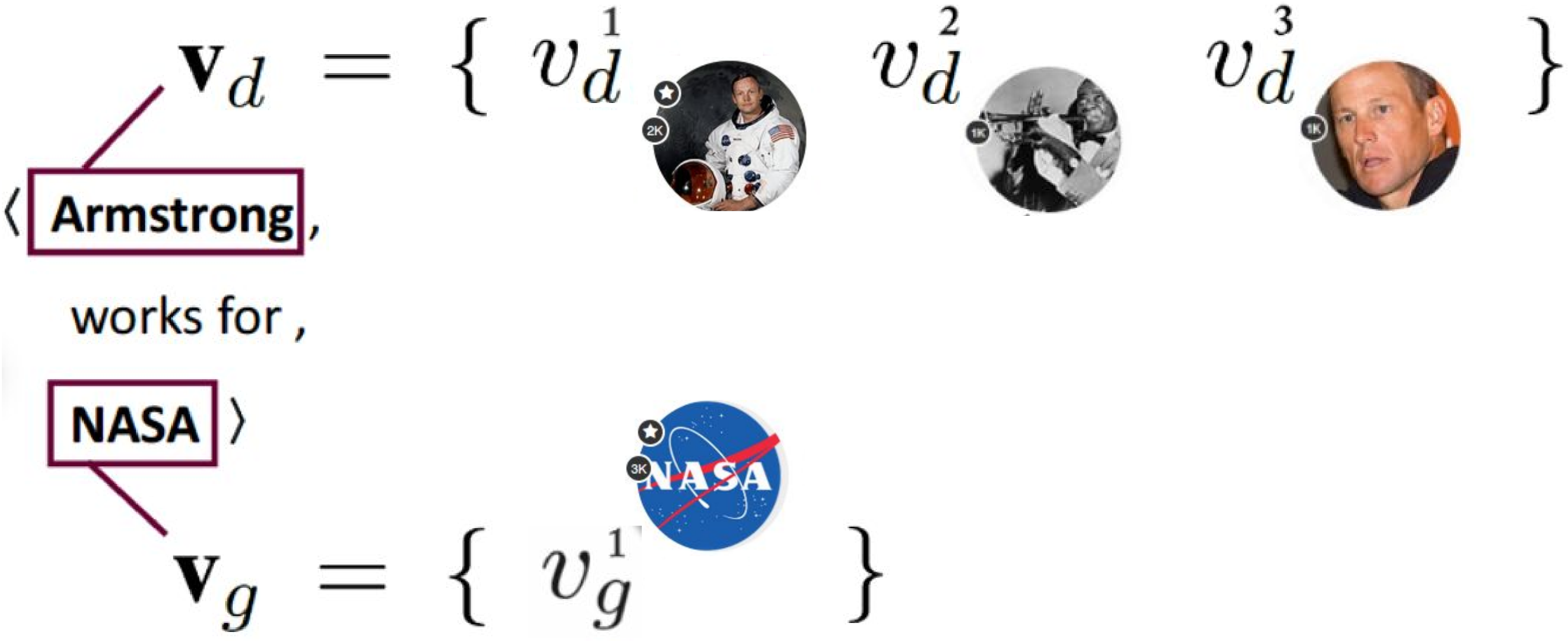
Group the set of unlinked triples by relation

For each relation **r**:

- Extract and disambiguate a subset of high-confidence **seed argument pairs** for **r** ;
- Estimate the **specificity** of **r** by looking at the distribution of its disambiguated seeds in the vector space $\mathbf{V_S}$;
- Disambiguate the remaining argument pairs of **r** with Babelfy either **triple-by-triple** (if **r** is general) or **all at once** (if **r** is specific).

KB-Unify: How it works

Identifying seed argument pairs



$$\mathbf{v}_d = \{ \; v_d^1 \qquad v_d^2 \qquad v_d^3 \; \}$$

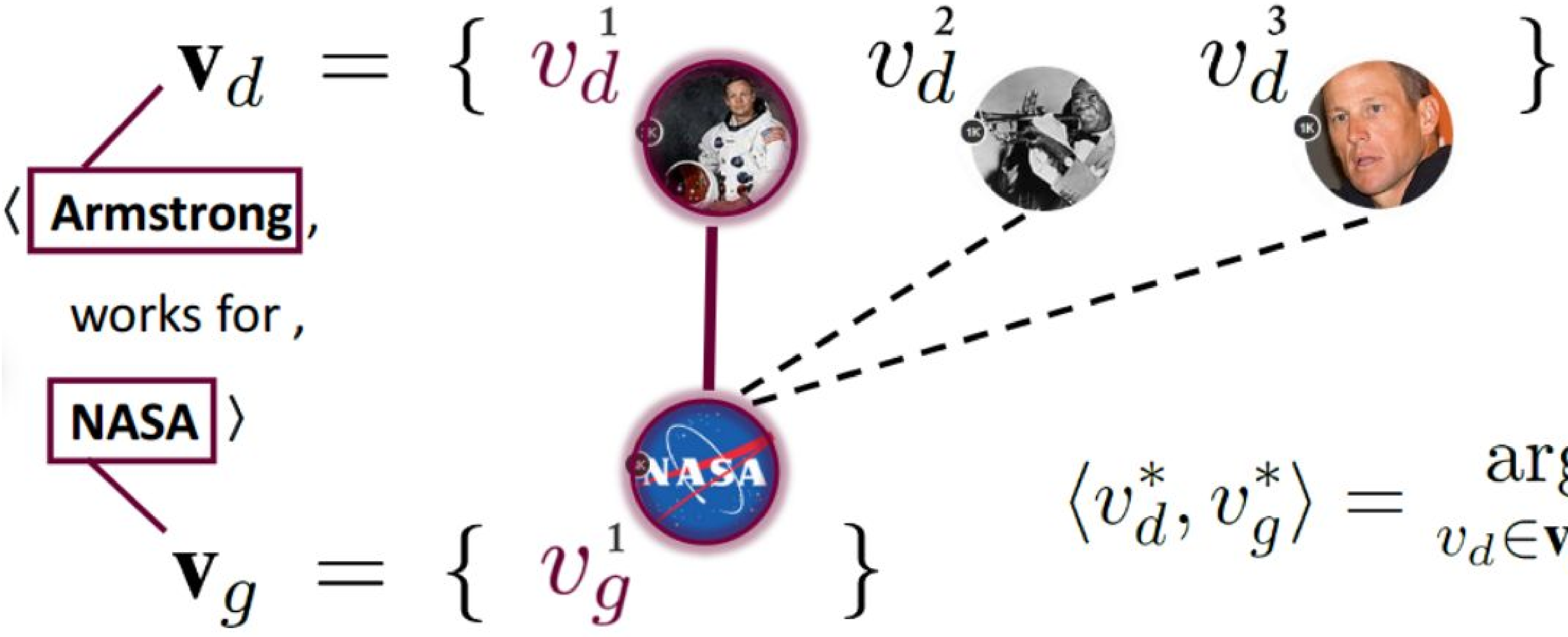
⟨ Armstrong , works for , NASA ⟩

$$\mathbf{v}_g = \{ \; v_g^1 \; \}$$

SENSEMBED

KB-Unify: How it works

Identifying seed argument pairs



$$\mathbf{v}_d = \{ \; v_d^1 \quad v_d^2 \quad v_d^3 \; \}$$

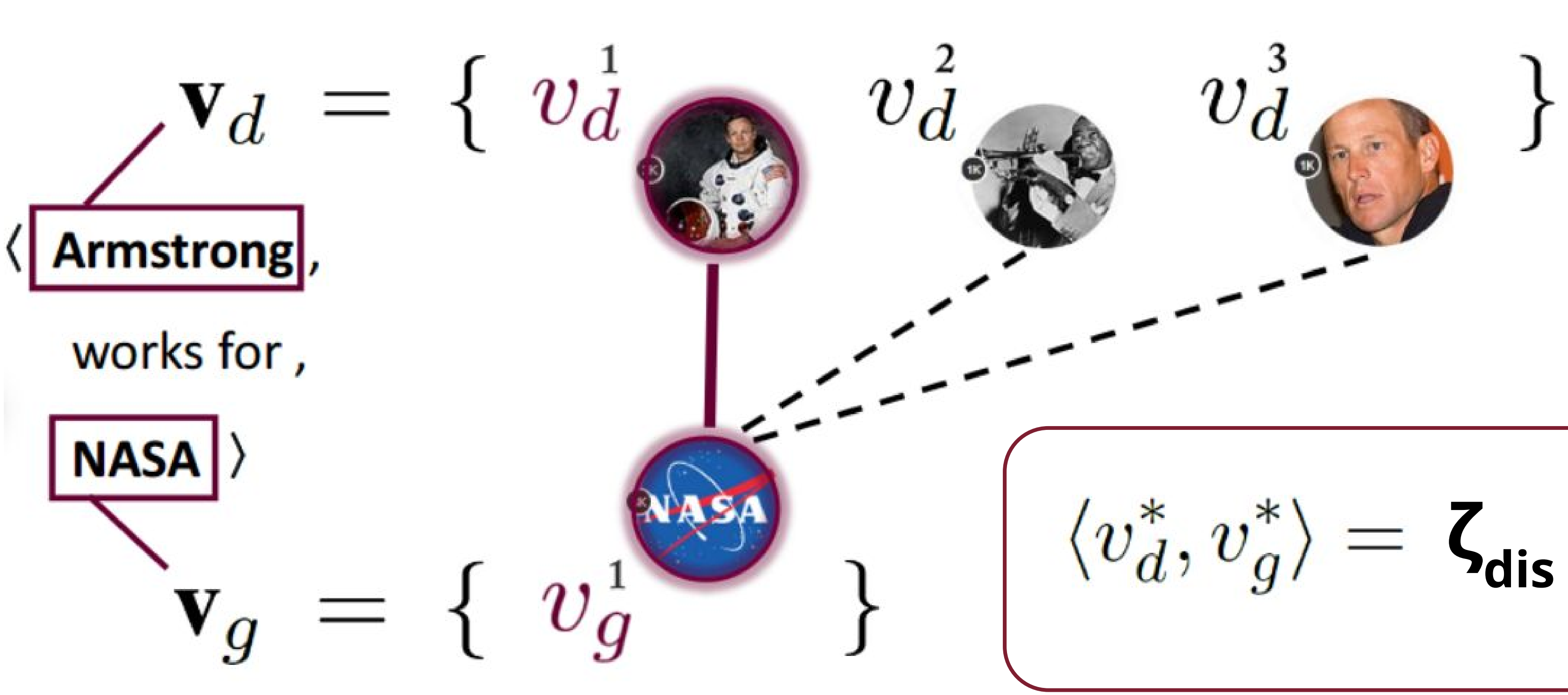
$$\mathbf{v}_g = \{ \; v_g^1 \; \}$$

SensEmbed

$$\langle v_d^*, v_g^* \rangle = \underset{v_d \in \mathbf{v}_d,\, v_g \in \mathbf{v}_g}{\operatorname{argmax}} \frac{v_d \cdot v_g}{\|v_d\| \, \|v_g\|}$$

KB-Unify: How it works

Identifying seed argument pairs



$$\mathbf{v}_d = \{\ v_d^1 \quad v_d^2 \quad v_d^3\ \}$$

⟨ Armstrong , works for , NASA ⟩

$$\mathbf{v}_g = \{\ v_g^1\ \}$$

SENSEMBED

$$\langle v_d^*, v_g^* \rangle = \zeta_{\text{dis}}$$

Seed Disambiguation Confidence

KB-Unify: How it works

Ranking relations by specificity

$$\mu_k = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} \frac{v}{||v||} \ , \quad k \in \{D, G\}$$

Domain/Range
Centroids

$$\sigma_k^2 = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} \left(1 - cos\left(v, \mu_k\right)\right)^2$$

Domain/Range
Variances

KB-Unify: How it works

Ranking relations by specificity

$$\mu_k = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} \frac{v}{||v||} \quad , \quad k \in \{D, G\}$$

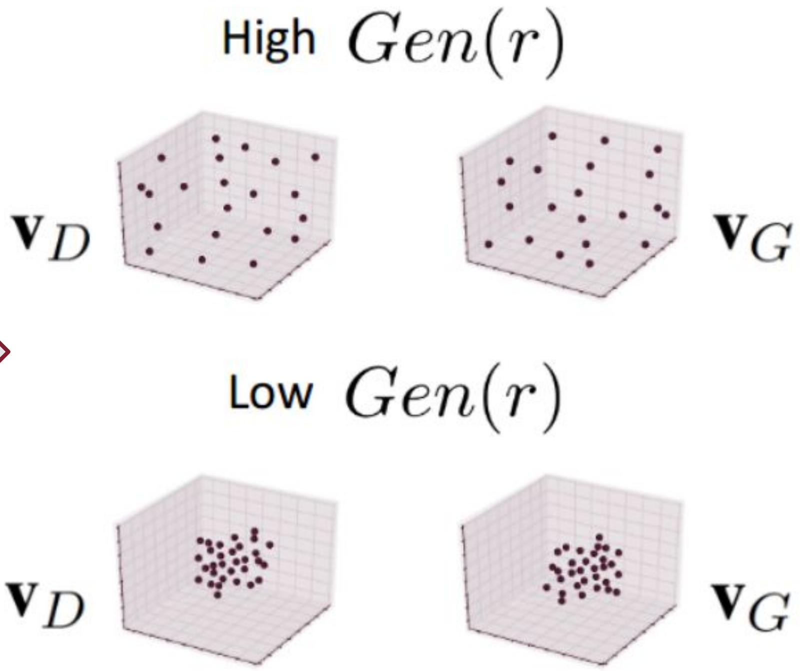
Domain/Range Centroids

$$\sigma_k^2 = \frac{1}{|\mathbf{v}_k|} \sum_{v \in \mathbf{v}_k} \left(1 - cos\left(v, \mu_k\right)\right)^2$$

Domain/Range Variances

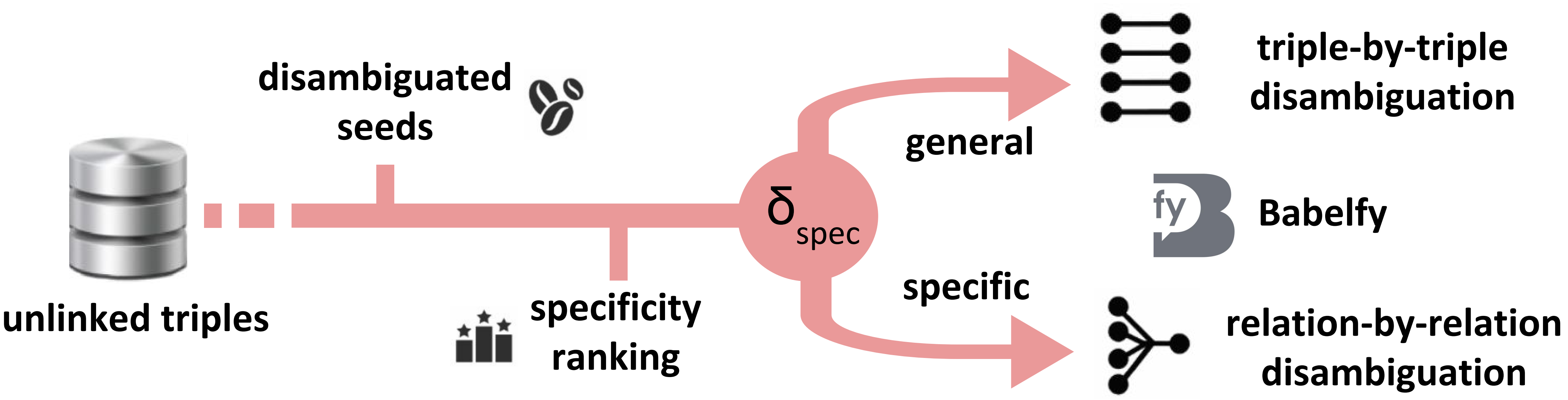
$$Gen(r) = \frac{\sigma_D^2 + \sigma_G^2}{2}$$

Specificity threshold: $\boldsymbol{\delta}_{spec}$



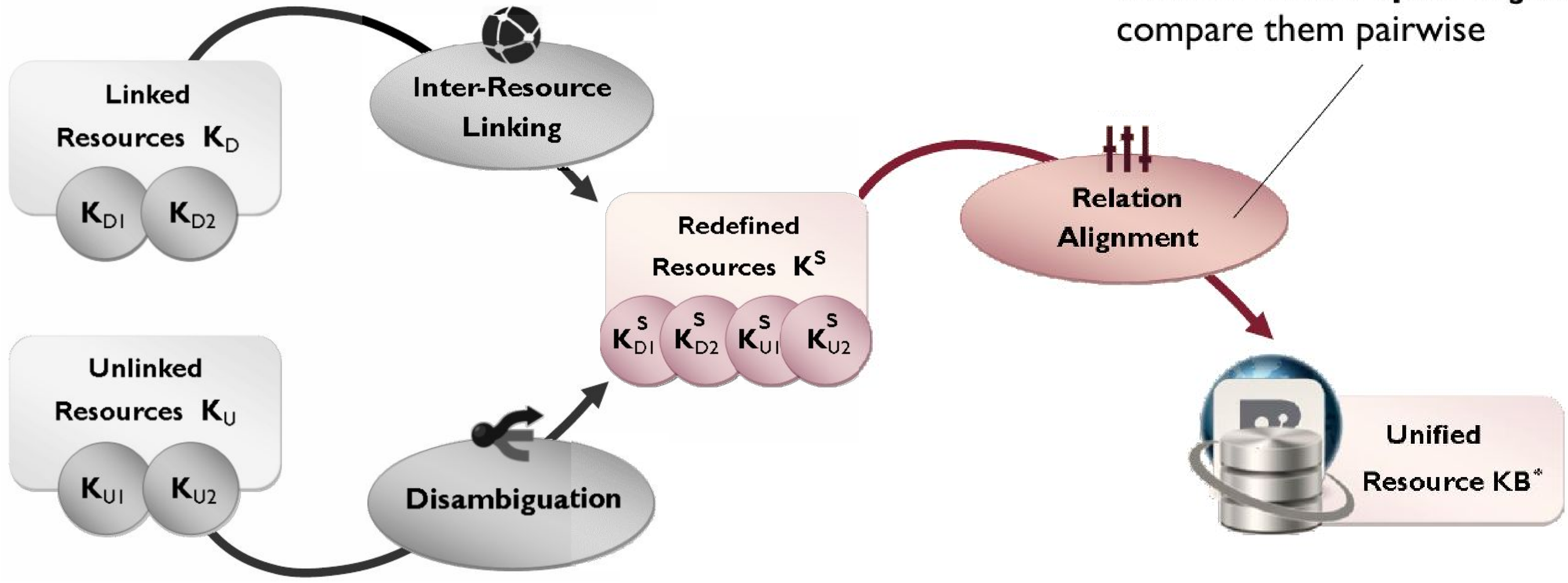
KB-Unify: How it works

Disambiguation with Relation Context



KB-Unify: How it works

A bird's-eye view



represent each relation in the **unified vector space V_S** and compare them pairwise

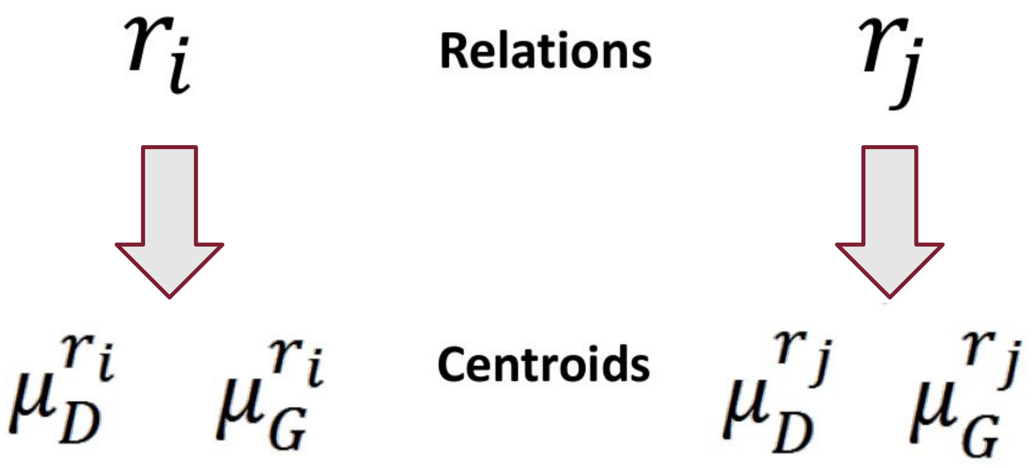
KB-Unify: How it works

Relation alignment

KB-Unify: How it works

Relation alignment

For each relation pair $\langle \boldsymbol{r_i}, \boldsymbol{r_j} \rangle$:



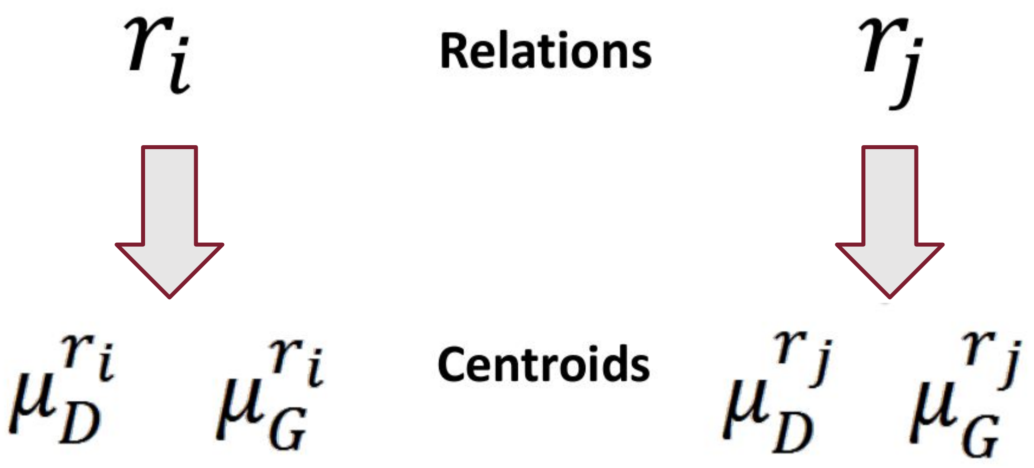
r_i **Relations** r_j

$\mu_D^{r_i}$ $\mu_G^{r_i}$ **Centroids** $\mu_D^{r_j}$ $\mu_G^{r_j}$

KB-Unify: How it works

Relation alignment

For each relation pair $\langle \boldsymbol{r_i}\,,\ \boldsymbol{r_j} \rangle$:



r_i **Relations** r_j

$\mu_D^{r_i}$ $\mu_G^{r_i}$ **Centroids** $\mu_D^{r_j}$ $\mu_G^{r_j}$

Compare domain and range centroids pairwise:

$$\underset{k\in\{D,G\}}{s_k} = \frac{\mu_k^{r_i}\cdot\mu_k^{r_j}}{\|\mu_k^{r_i}\|\;\|\mu_k^{r_j}\|}$$

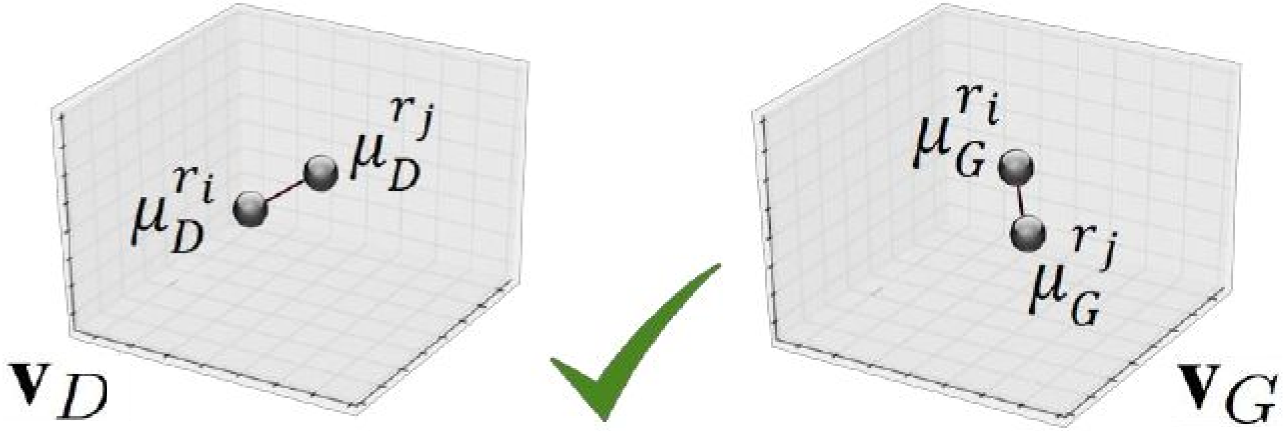
Relation Centroid Similarity

KB-Unify: How it works

Relation alignment

Fix a similarity threshold δ_{align}:

Domain Centroids



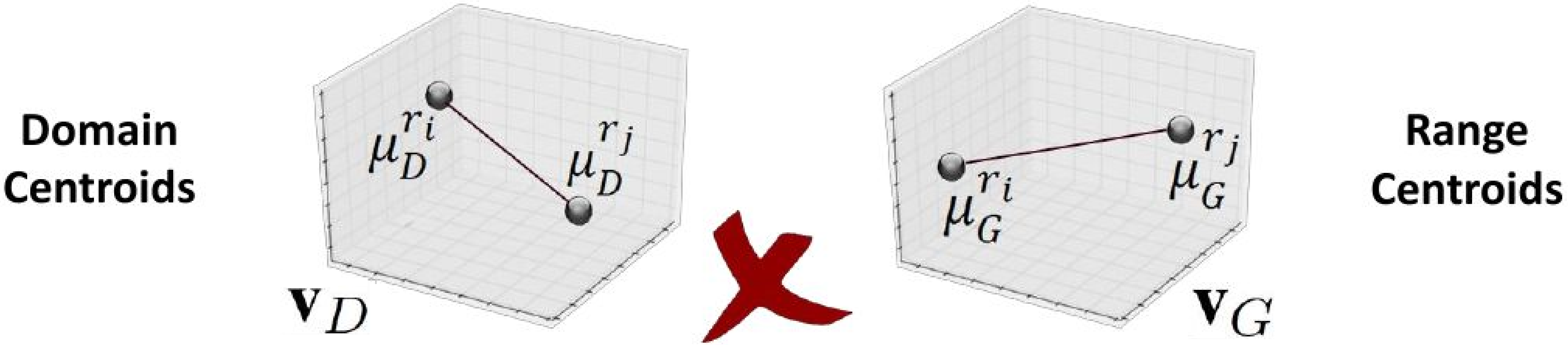
Range Centroids

$\frac{1}{2}(\boldsymbol{s_D} + \boldsymbol{s_G}) \geq \boldsymbol{\delta_{align}}$? Align $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$ and merge them in the same cluster

KB-Unify: How it works

Relation alignment

Fix a similarity threshold δ_{align}:

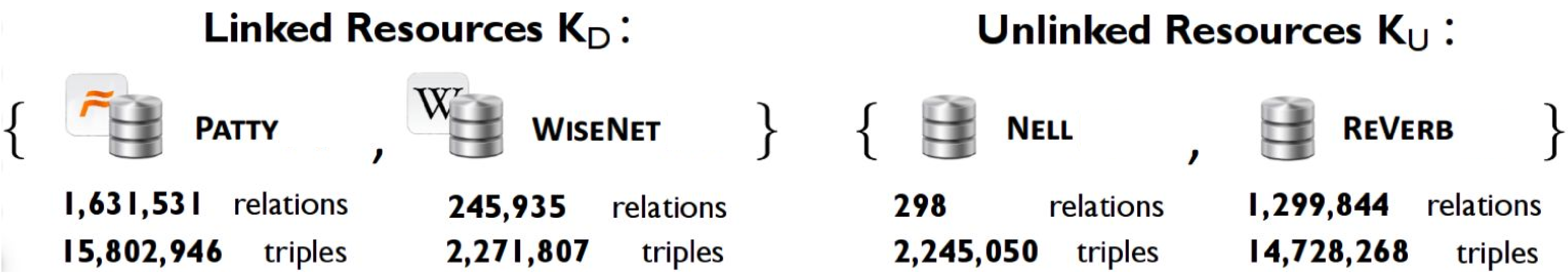


$\frac{1}{2}(\boldsymbol{s_D} + \boldsymbol{s_G}) < \boldsymbol{\delta_{align}}$? Leave $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$ in separate clusters

KB-Unify: Experiments

Evaluation

Experimental setup:



Linked Resources K_D : { PATTY , WISENET }

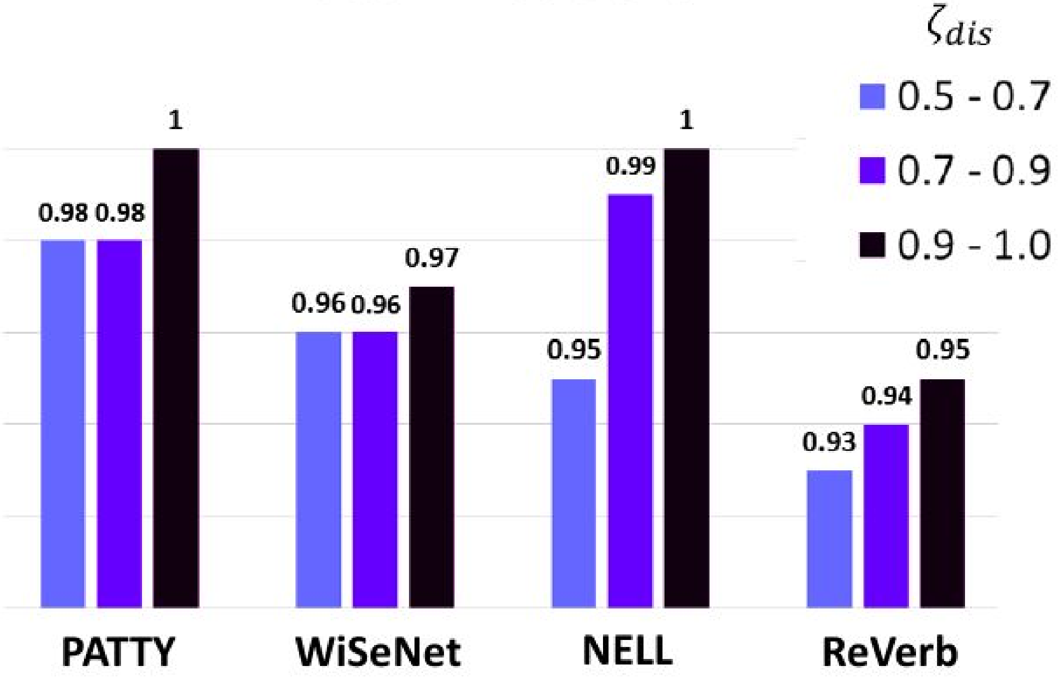
Unlinked Resources K_U : { NELL , REVERB }

	PATTY	WISENET	NELL	REVERB
relations	1,631,531	245,935	298	1,299,844
triples	15,802,946	2,271,807	2,245,050	14,728,268

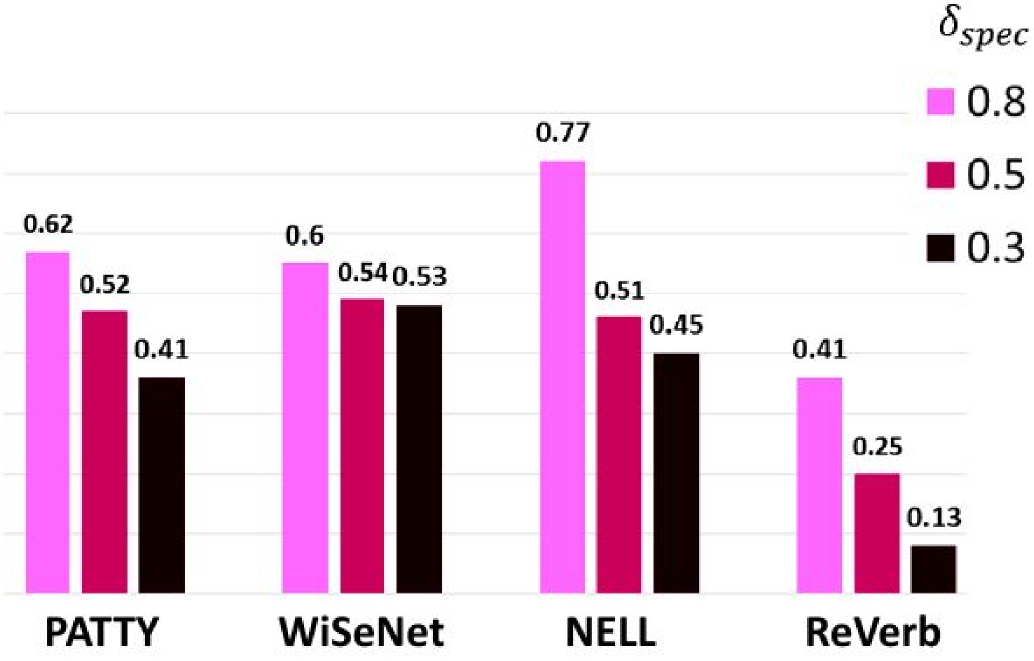
KB-Unify: Experiments

Disambiguation

Seed Precision:



Coverage:



KB-Unify: Experiments

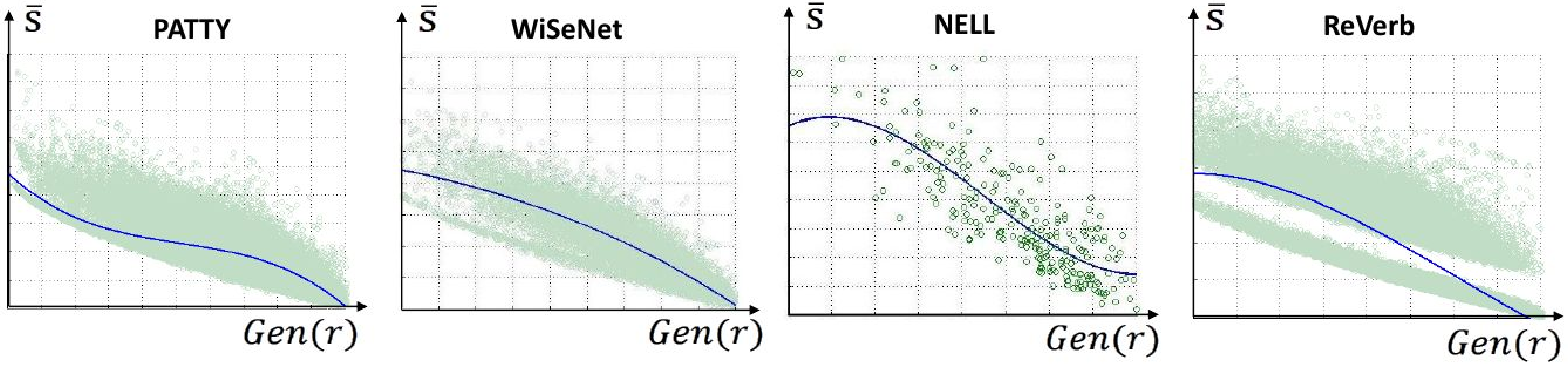
Specificity ranking

For each ranked relation compute $Gen(r)$ against the **average argument similarity** $\bar{s}$:

KB-Unify: Experiments

Specificity ranking

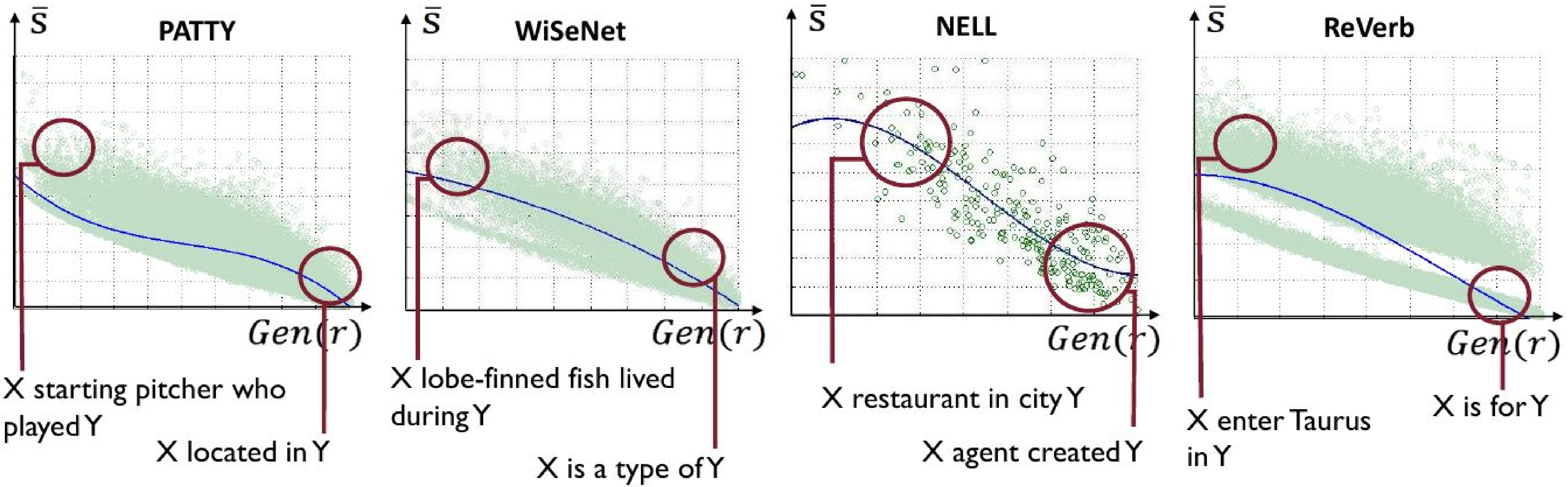
For each ranked relation compute $Gen(r)$ against the **average argument similarity** $\bar{s}$:



KB-Unify: Experiments

Specificity ranking

For each ranked relation compute $Gen(r)$ against the **average argument similarity $\bar{s}$**:



X starting pitcher who played Y

X located in Y

X lobe-finned fish lived during Y

X is a type of Y

X restaurant in city Y

X agent created Y

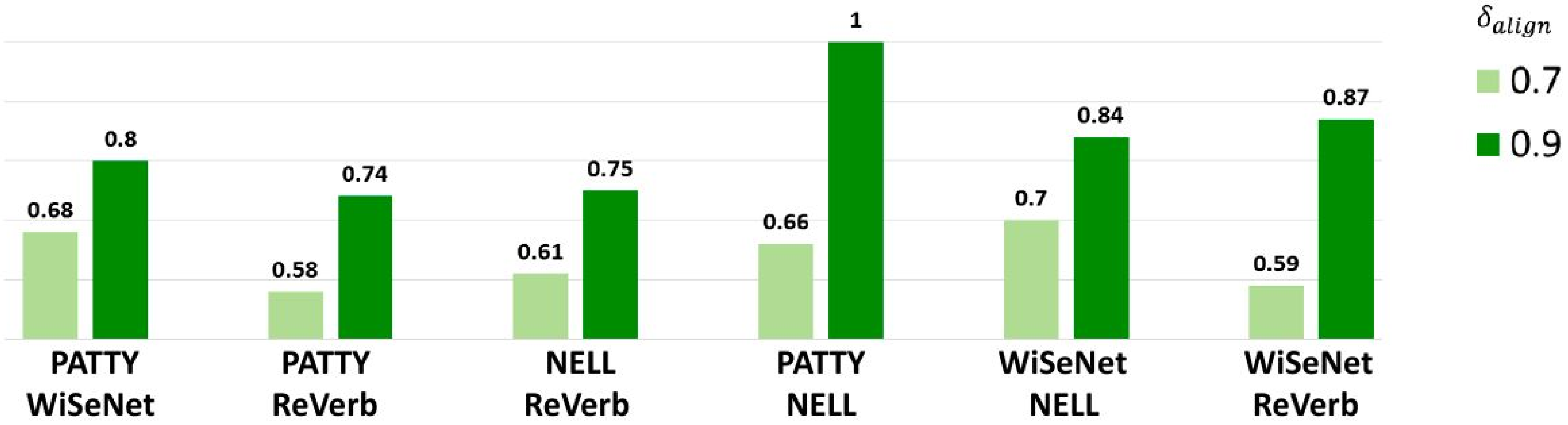
X enter Taurus in Y

X is for Y

KB-Unify: Experiments

Cross-resource relation alignment

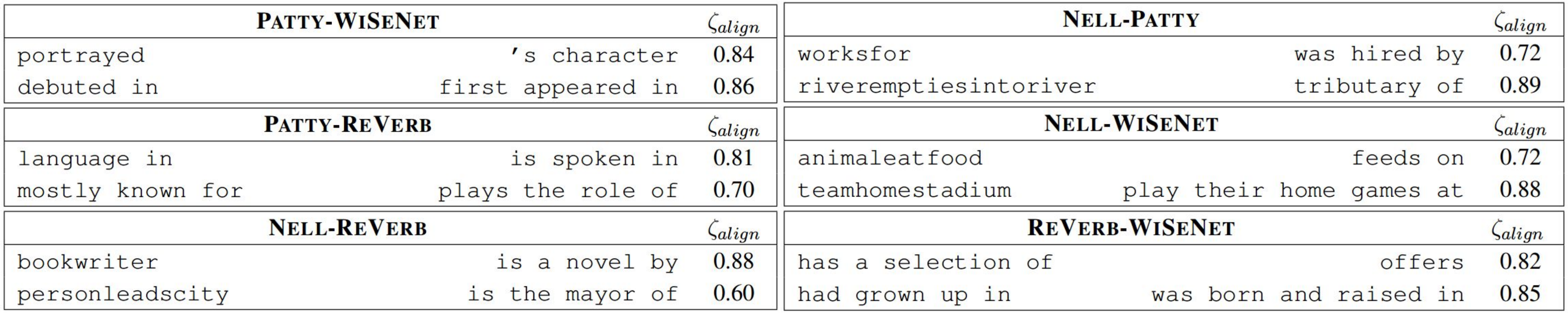
Samples of **150 candidate alignments** for different alignment thresholds δ_{align} manually evaluated (in terms of **paraphrasing**) by two human judges



KB-Unify: Experiments

Cross-resource relation alignment

Some examples:



PATTY-WISENET		ζ_{align}
portrayed	's character	0.84
debuted in	first appeared in	0.86

NELL-PATTY		ζ_{align}
worksfor	was hired by	0.72
riveremptiesintoriver	tributary of	0.89

PATTY-REVERB		ζ_{align}
language in	is spoken in	0.81
mostly known for	plays the role of	0.70

NELL-WISENET		ζ_{align}
animaleatfood	feeds on	0.72
teamhomestadium	play their home games at	0.88

NELL-REVERB		ζ_{align}
bookwriter	is a novel by	0.88
personleadscity	is the mayor of	0.60

REVERB-WISENET		ζ_{align}
has a selection of	offers	0.82
had grown up in	was born and raised in	0.85

KB-Unify: Future work

Where from here?

- Less “naïve” **relation alignment** procedure
- **Iterative** algorithm for disambiguation and alignment (EM-style)
- Unify OIE-based KBs with **hand-curated resources** (Wikidata, DBpedia, etc.)

...

Wrap up and Conclusion

Wrap up and Conclusion

DefIE: A full-fledged OIE pipeline targeted to textual definitions, with explicit semantic characterization of both arguments and relation patterns

Wrap up and Conclusion

DefIE: A full-fledged OIE pipeline targeted to textual definitions, with explicit semantic characterization of both arguments and relation patterns

KB-Unify: An approach to knowledge base disambiguation and unification based on a shared sense inventory and a sense-based vector space model

Wrap up and Conclusion

Take-home message(s):

Web-scale OIE is absolutely great, but…

Wrap up and Conclusion

Take-home message(s):

Web-scale OIE is absolutely great, but…

1. **Definitional knowledge is important**: sometimes it is worth just stepping back and analyze from where valuable information is extracted (**quality vs. quantity**)

Wrap up and Conclusion

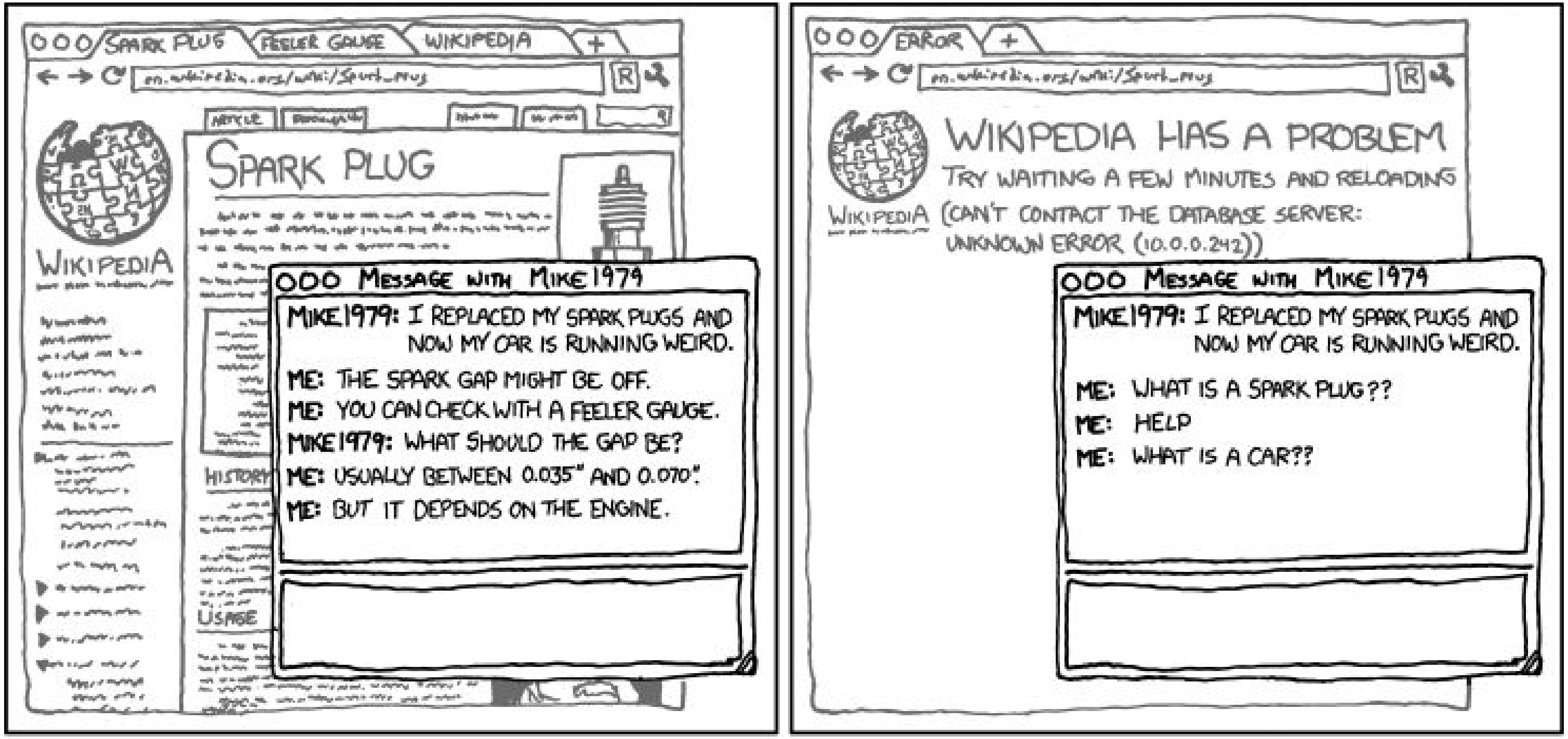
Take-home message(s):

Web-scale OIE is absolutely great, but...

1. **Definitional knowledge is important**: sometimes it is worth just stepping back and analyze from where valuable information is extracted (**quality vs. quantity**)

2. **Making sense of the output is important:** semantic analysis can be used to let different OIE outputs "speak to each other" and benefit from mutual enrichment

Thank you!


SPARK PLUG
FEELER GAUGE
WIKIPEDIA
SPARK PLUG
WIKIPEDIA
MESSAGE WITH MIKE1979
MIKE1979: I REPLACED MY SPARK PLUGS AND NOW MY CAR IS RUNNING WEIRD.
ME: THE SPARK GAP MIGHT BE OFF.
ME: YOU CAN CHECK WITH A FEELER GAUGE.
MIKE1979: WHAT SHOULD THE GAP BE?
ME: USUALLY BETWEEN 0.035" AND 0.070".
ME: BUT IT DEPENDS ON THE ENGINE.
HISTORY
USAGE
ERROR
WIKIPEDIA HAS A PROBLEM
TRY WAITING A FEW MINUTES AND RELOADING
WIKIPEDIA
(CAN'T CONTACT THE DATABASE SERVER: UNKNOWN ERROR (10.0.0.242))
MESSAGE WITH MIKE1979
MIKE1979: I REPLACED MY SPARK PLUGS AND NOW MY CAR IS RUNNING WEIRD.
ME: WHAT IS A SPARK PLUG??
ME: HELP
ME: WHAT IS A CAR??

WHEN WIKIPEDIA HAS A SERVER OUTAGE, MY APPARENT IQ DROPS BY ABOUT 30 POINTS.

xkcd, “Extended Mind”