

---

# Extending a Large Semantic Network

---

Luis Espinosa-Anke

# Extending a Large Semantic Network

---

- COST STSM Reference Number: COST-STSM-IS1305- 160215-055460
- Host Institution/Researcher: Sapienza University of Rome / Prof. Roberto Navigli

## Talk outline

---

- STSM Motivation
- Knowledge Base Unification via Sense-based Embeddings and Disambiguation
- Discussion: Interaction with Lexicography

## STSM Motivation

---

- BabelNet (Navigli and Ponzetto, 2012) is currently the largest repository for entities and concepts, containing up to 14M BNsynsets.
- It is fully taxonomized thanks to the Wikipedia Bitaxonomy, i.e. WP pages and categories (Flati et al., 2014).

# STSM Motivation



BabelNet

LOG IN REGISTER

European Cooperation in Science and Technology

ENGLISH

TRANSLATE INTO...

SEARCH

PREFERENCES

English

Arabic

Chinese

French

German

Greek

Hebrew

Hindi

Italian

+ all preferred languages

Dictionary

Images

Translations

Sources

Categories

External links

bn:00546233n • NOUN • Named Entity • Categories: Intergovernmental organizations, Science and technology in Europe, Funding bodies, 1971 establishments in Belgium...

**European Cooperation in Science and Technology · COST**   · **COST - European Cooperation in the Field of Scientific and Technical Research · COST – European Cooperation in the Field of Scientific and Technical Research · COST – European Cooperation in Science and Technology · COST - European Cooperation in Science and Technology**

European Cooperation in Science and Technology is Europe's longest-running intergovernmental framework for cooperation in science and technology.  *More definitions*

# STSM Motivation



BabelNet

[LOG IN](#) [REGISTER](#)

European Cooperation in Science and Technology

ENGLISH

TRANSLATE INTO...

SEARCH

[PREFERENCES](#)

English

Arabic

Chinese

French

German

Greek

Hebrew

Hindi

Italian

[+](#) all preferred languages

● Dictionary

● Images

● Translations

● Sources

● Categories

● External links

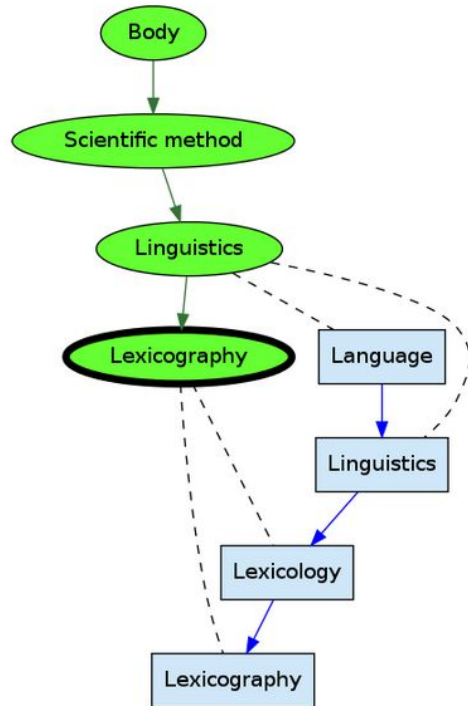
bn:00546233n · NOUN · Named Entity · Categories: Intergovernmental organizations, Science and technology in Europe, Funding bodies, 1971 establishments in Belgium...

**🇬🇧 European Cooperation in Science and Technology · COST**   · **COST - European Cooperation in the Field of Scientific and Technical Research · COST – European Cooperation in the Field of Scientific and Technical Research · COST – European Cooperation in Science and Technology · COST - European Cooperation in Science and Technology**

 Wikipedia categories

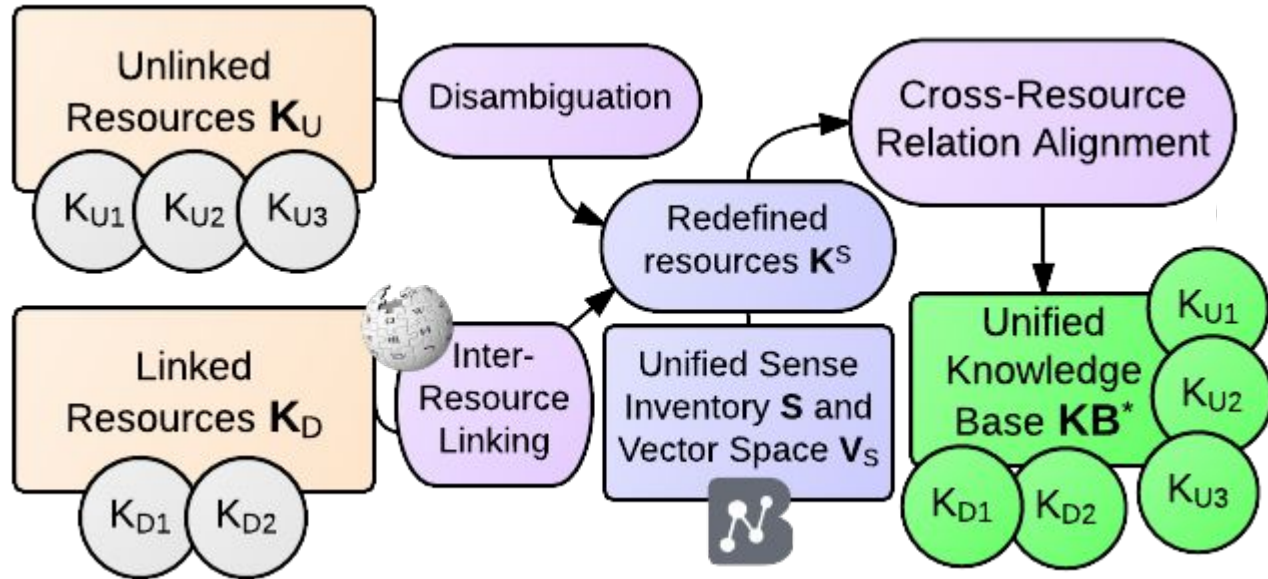
**🇬🇧** [Intergovernmental organizations, Science and technology in Europe, Funding bodies, 1971 establishments in Belgium, Organizations established in 1971, 1971 in science, Organisations based in Brussels, Science and technology in Belgium](#)

# STSM Motivation



- Aiming at introducing finer-grained relations into BabelNet.
- Providing an ontological structure to the BabelNet graph.
- Leverage existing knowledge derived from successful OIE systems: NELL (Carlson et al., 2010), PATTY (Nakashole et al., 2012), WiseNet (Moro and Navigli, 2012) and ReVerb (Etzioni et al., 2011).

# Knowledge Base Unification - Summary





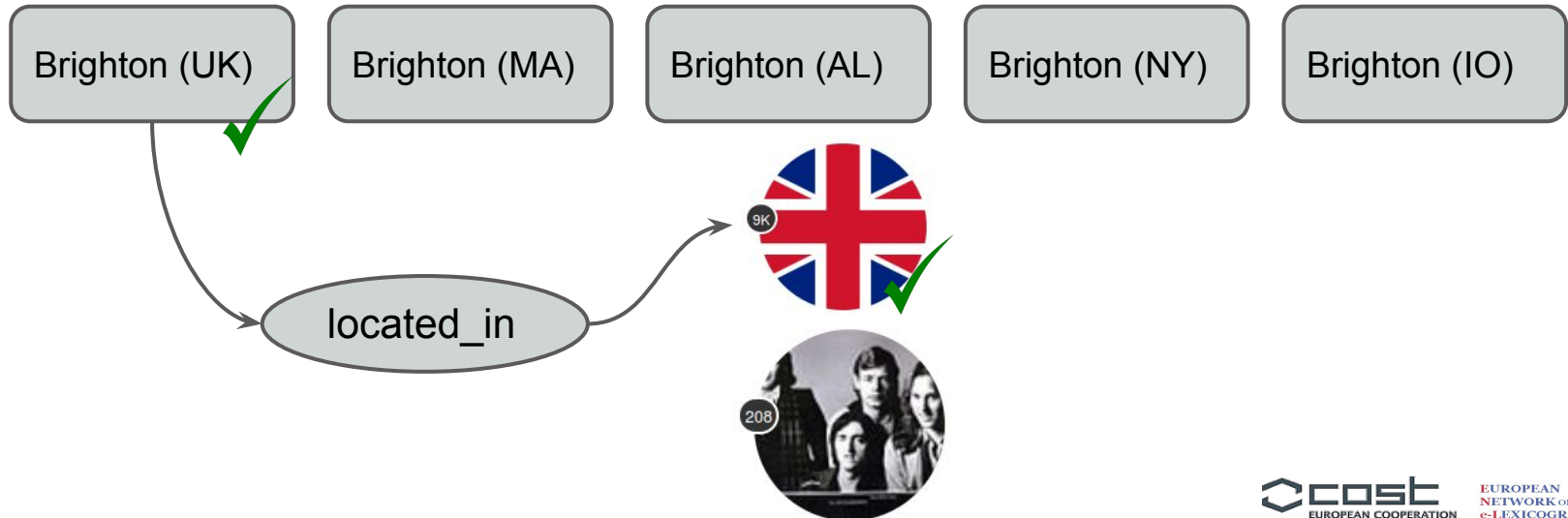
# Knowledge Base Unification - (1) Disambiguation

---

- Brighton is located in the UK
- Rooney plays for The Red Devils
- NLP is a subfield of AI
- ...

# Knowledge Base Unification - (1) Disambiguation

- Brighton is located in the UK



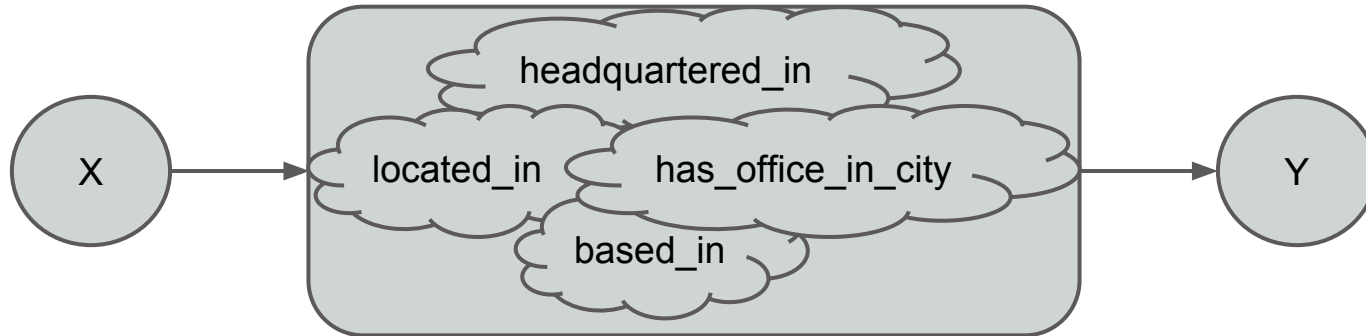
## Knowledge Base Unification - (2) Relation Alignment

---

- X is located in Y
- X is headquartered in Y
- X is based in Y
- X has office in city Y
- ...

# Knowledge Base Unification - (2) Relation Alignment

---



# Knowledge Base Unification - Resource Statistics

---

- Still haven't decided on release details
- Currently:
  - .95 alignment confidence:
    - 1105 relation synsets
    - Avg. Size 6.11
    - Not a single *singleton* synset.



## Reference










---

- System description, experiments and evaluation details:

Delli Bovi, C., Espinosa-Anke, L. and Navigli, R. (2015). **Knowledge Base Unification via Sense-based Embeddings and Disambiguation**. In *Proceedings of EMNLP 2015*. Lisbon, Portugal.

# Thanks!

---

-  感謝, 感谢, 谢谢
-  thank you
-  remerciement, *merci, je vous remercie*
-  Danksagung, Dankbezeugung, *danke*
-  ευχαριστώ, ευχαριστήρια, *σας ευχαριστώ*
-  धन्यवाद, शुक्रिया
-  grazie, ringraziamento
-  ありがとう, サンキュー
-  спасибо
-  gracias, agradecimiento, *¡gracias*

## Q&A?

---

- Use disambiguated relation triples for enriching existing lexicographic resources?
- As a knowledge benchmark to assess fitness of dictionary examples?
- As a provider for hierarchical structure of domain-specific glossaries or dictionaries?