



Semantic technologies for Digital Libraries

Andrea Giovanni Nuzzolese

¹STLab, Institute of Cognitive Science and Technology, National Research Council, Italy



Outline

- Semantic enhancement of content for DLs
 - Entity resolution and Linking, i.e., Apache Stanbol Enhancer
- Machine-reading based knowledge extraction
 - STLab tools
- Semantic search
 - Entity-centric indexing



Entity Resolution and Linking

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”



Entity Resolution and Linking

“**Iran** has tentatively agreed to ship much of its huge stockpile of **uranium** to **Russia** if it reaches a broader **nuclear** deal with the **West**...”



Entity Resolution and Linking

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”

geonames:Iran

dbpedia:Uranium

geonames:Russia

dbpedia:Western_world

dbpedia:Nuclear_power

Entity Resolution and Linking

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”

geonames:Iran

geonames:Russia

dbpedia:Nuclear_power

dbpedia:Uranium

dbpedia:Western_world



- capital: dbpedia:Teheran
- government type: dbpedia:Islamic_republic
- leader: dbpedia:Hassan_Rouhani
- ...

Apache Stanbol Enhancer

- The Enhancer component enables entity recognition and linking to (any) existing dataset in Linked Data
 - natural language processing and metadata extraction
- It provides a machinery to further process this data and add additional knowledge and links via applying rules and reasoning.
- It is built as a modular set of components that can be combined according to specific tasks
- The Enhancer provides both a RESTful and a Java API
- Apache Stanbol results from IKS (EU FP7 IP) in which the STLab was a research partner



Apache Stanbol Enhancer in a nutshell

```
curl -X POST -H "Accept: text/turtle" -H "Content-type: text/plain" \  
  --data "The Stanbol enhancer can detect famous cities such as \  
    Paris and people such as Bob Marley." \  
  http://localhost:8080/enhancer
```



Enhancement Chain: **default** all 5 engines available

- 🔌 **tika** (optional, TikaEngine)
- 🔌 **langid** (required, LangidEnhancementEngine)
- 🔌 **nerf** (required, NamedEntityExtractionEnhancementEngine)
- 🔌 **dbpediaLinking** (required, NamedEntityTaggingEngine)



Extracted entities

People

Places

```
{  
  "dc:creator": "Bob Marley",  
  "dc:relation": "urn:stanbol.org:stanbol-core:entity-referent",  
  "enhancer:confidence": 17396.67,  
  "enhancer:entity-label": {  
    "@literal": "Bob Marley",  
    "@language": "en"  
  },  
  "enhancer:entity-reference": "http://dbpedia.org/resource/Bob_Marley",  
  "enhancer:entity-type": [  
    "dbp-ont:MusicalArtist",  
    "foaf:Person",  
    "dbp-ont:Artist",  
    "dbp-ont:Person",  
    "owl:Thing"  
  ]  
}
```

RDF



Machine-reading based knowledge extraction with SHELDON



SHELDON
Powered by **STLAB**

RESTful

YouTube

Type a text | Upload a document

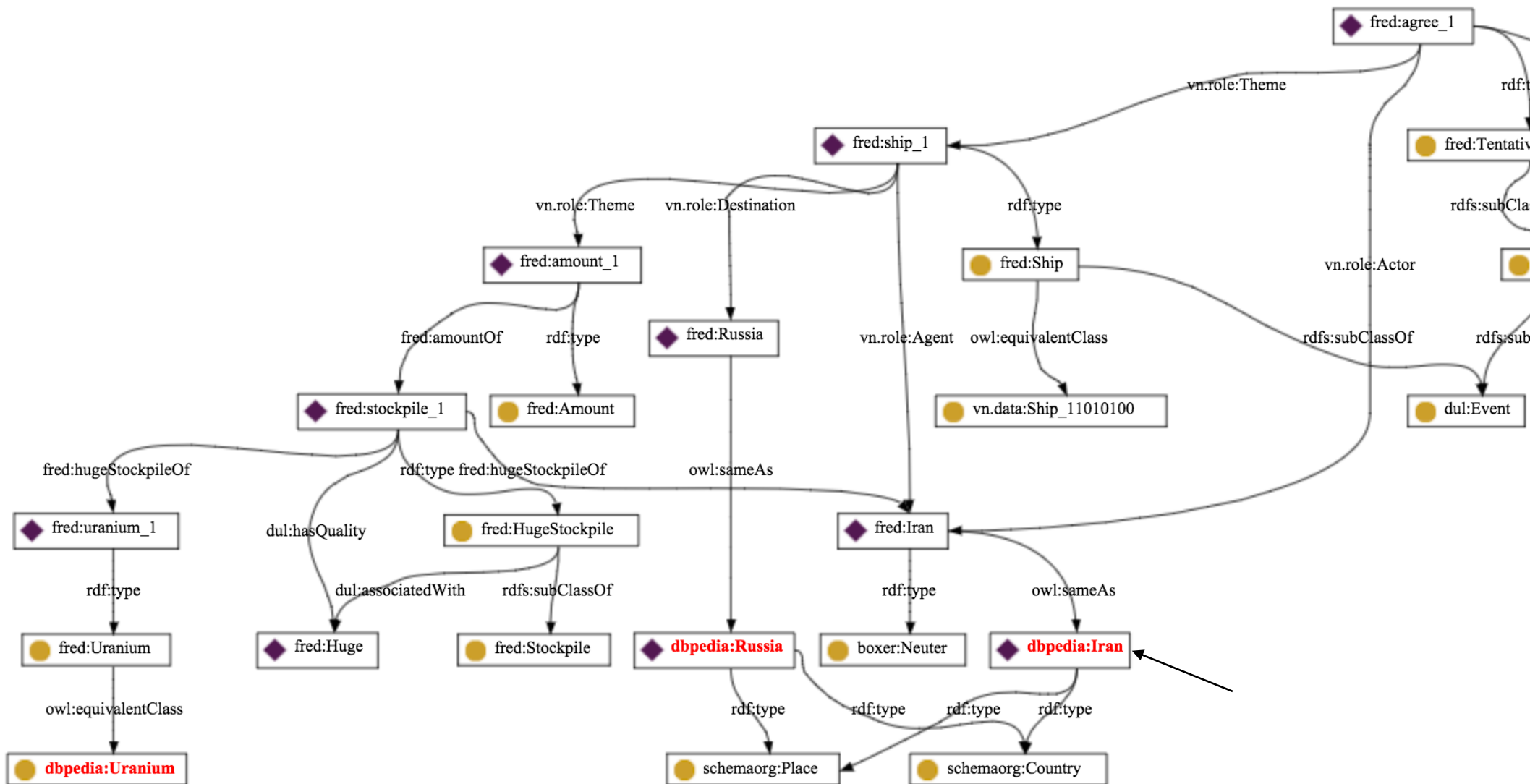
Produce output as: HTML

56 Ba Barium 137.327	247 Zr Zirconium 300	53 I Iodine 126.90447	7 N Nitrogen 14.00674	31 Ga Gallium 69.723
--------------------------------------	--------------------------------------	---------------------------------------	---------------------------------------	--------------------------------------

<http://wit.istc.cnr.it/stlab-tools/sheldon>

FRED

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”



FRED

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”

Semantic roles

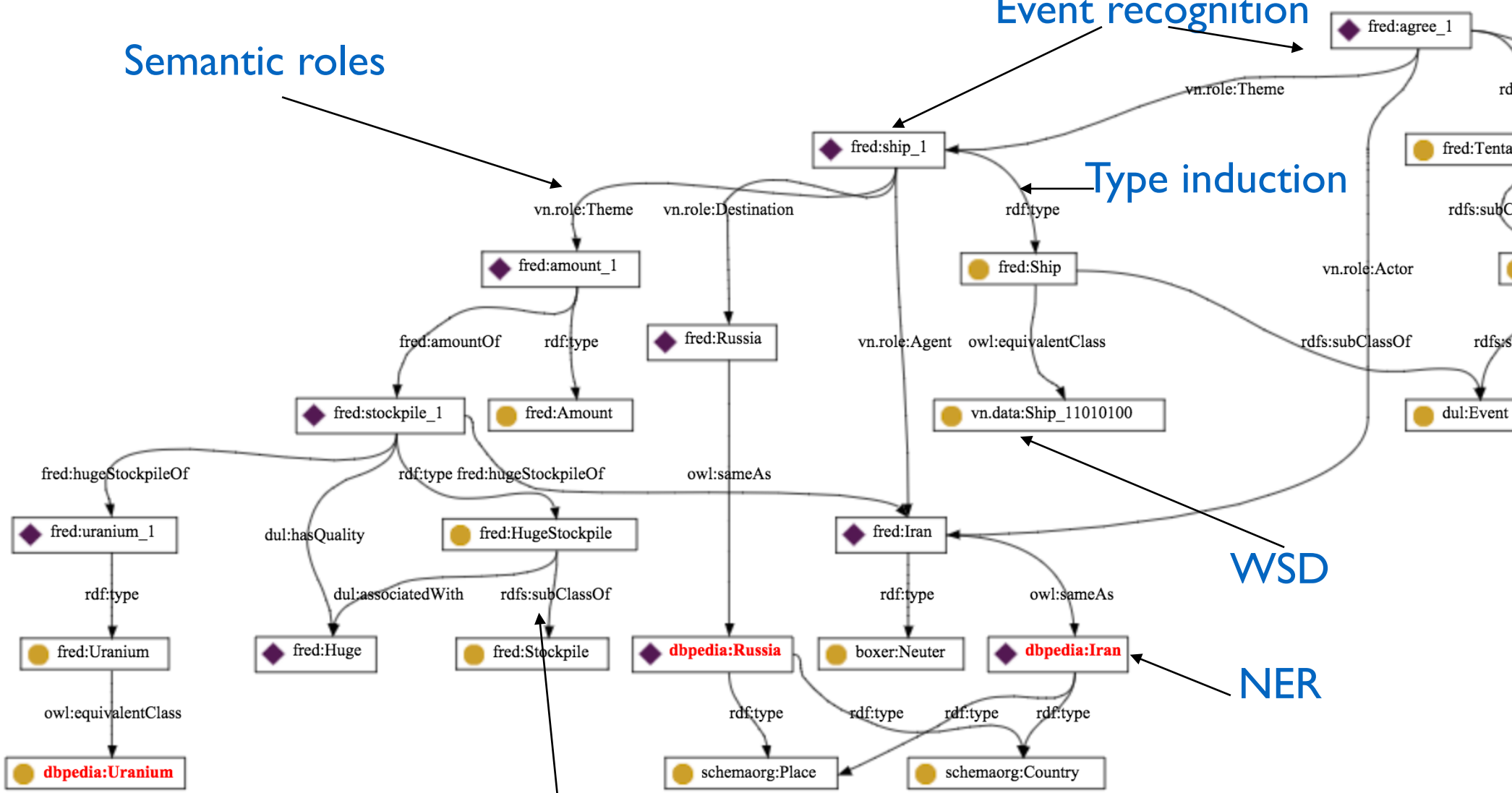
Event recognition

Type induction

WSD

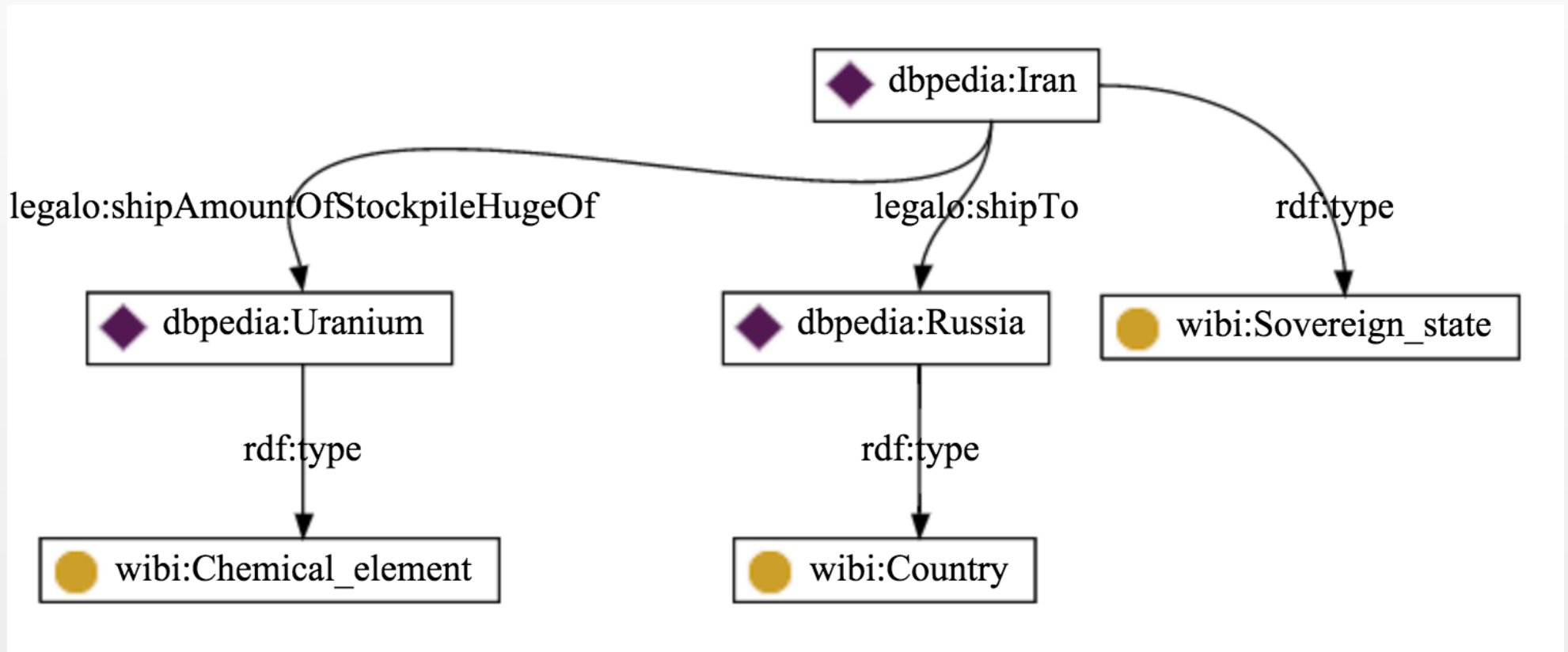
NER

Taxonomy induction



Relation discovery with Legalo

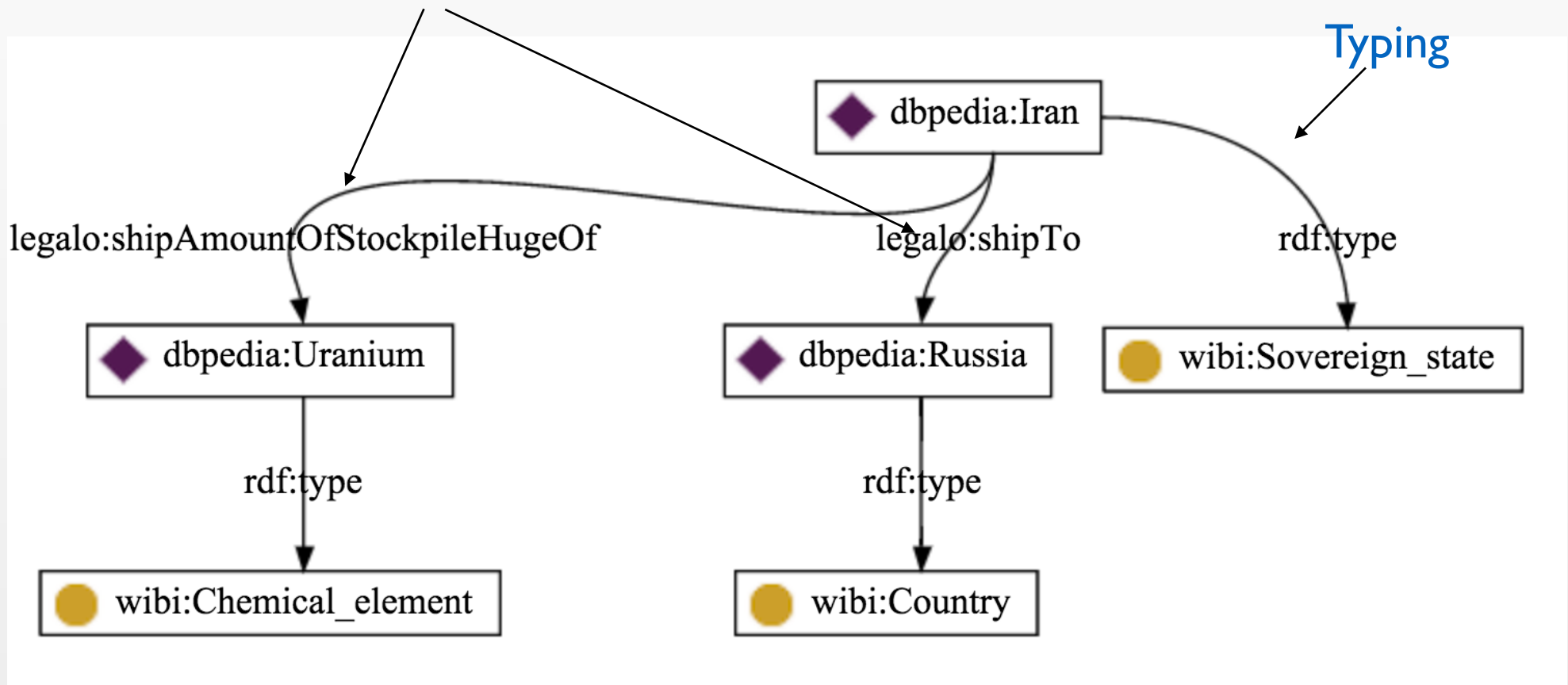
“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”



Relation discovery with Legalo

“Iran has tentatively agreed to ship much of its huge stockpile of uranium to Russia if it reaches a broader nuclear deal with the West...”

Entity relations from natural language



Citation typing with CiTalO

- Motivations
 - Bibliographic citations can be seen as tools for linking, disseminating, exploring and evaluating research and scholarly articles
 - These activities can be radically improved by exploiting the actual function of citations, i.e. author's reason for citing a given paper
- Goal
 - To design a method able to automatically infer the author's reason for citing a scientific article



Citation typing with CiTalO

“... For this, I highly recommend X. ...”



Citation typing with CiTalO

“... For this, I highly recommend X. ...”

The citing entity cites the cited entity as an item of recommended reading. This property can be used, for example, to describe references in a lecture reading list, where the cited references are relevant to the general topic of the lecture, but might not be individually cited within the text of the lecture. Similarly, it could be used to describe items in a 'Suggested further reading' list at the end of a book chapter.

y the following CiTO

- [cito:citesAsRecommendedReading](#)



Semantic search

- Entity-centric indexing Vs. token indexing
 - i.e., semiotic indexing of content based on the associations of words, entities and senses
- The search is performed by looking-up the semiotic indexing
- Additional information for a set of given entities can be meshed-up from Linked Data according to specific tasks
 - e.g., images, labels, owl:sameAs, etc.

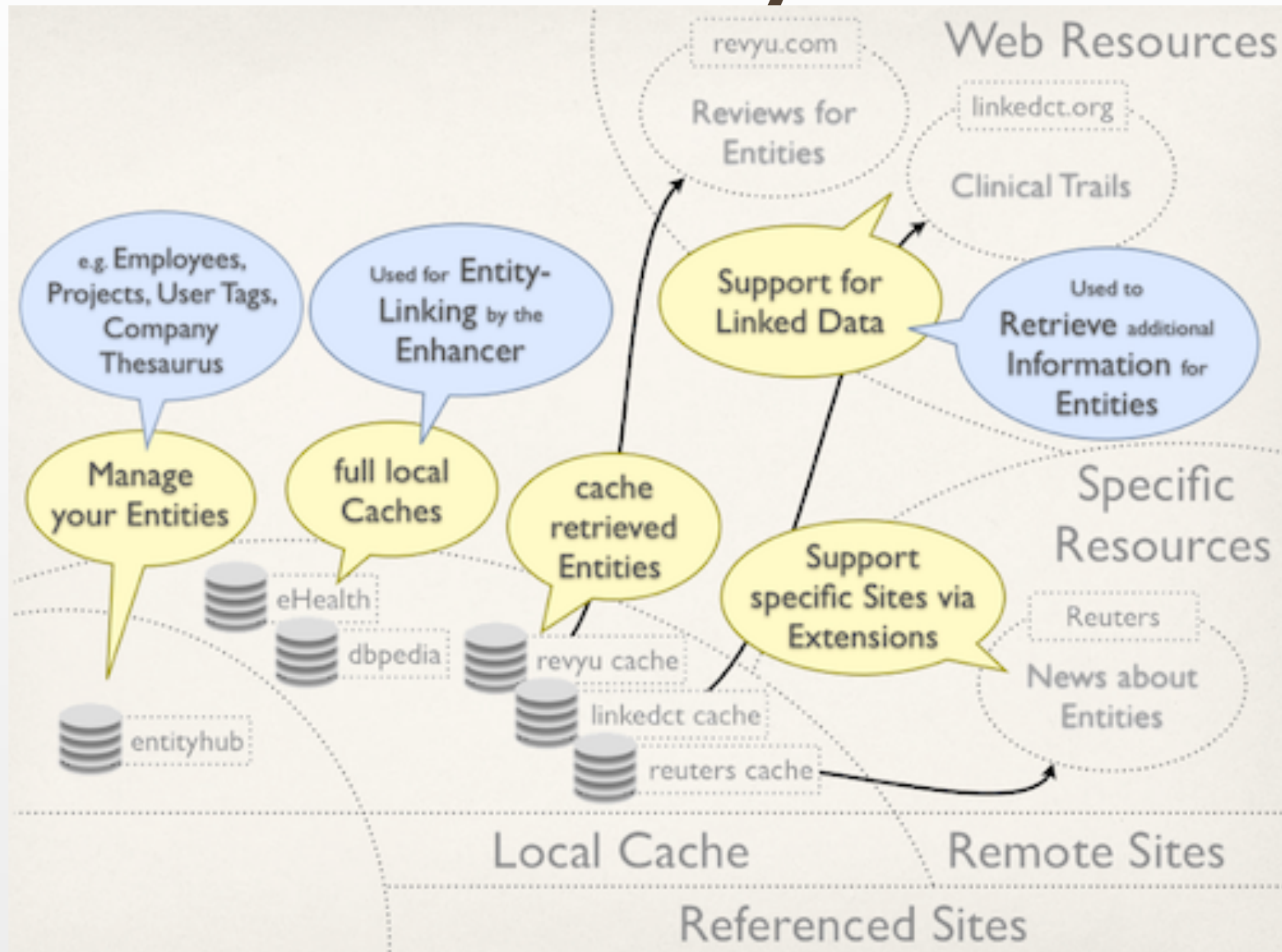


Apache Stanbol Entityhub

- The Entityhub is a component of Apache Stanbol that is compliant with the semiotic indexing approach
- It is responsible for
 - Indexing of a network of remote sites for fast entity lookup
 - Caching of externally retrieved entity information
 - CRUD management of local entities
- It exposes its functionality through HTTP RESTful services



Apache Stanbol Entityhub: overview



Thank you!

