

# Intermediate SPARQL Querying II

LiLa Corpora

Francesco Mambrini

4th Summer Datathon on Linguistic Linked Open Data  
Cercedilla, Spain  
1st June 2022



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.

- ▶ [lila-erc.eu/sparql/](http://lila-erc.eu/sparql/)
- ▶ [github.com/CIRCSE/SPARQL-queries](https://github.com/CIRCSE/SPARQL-queries)

## Introduction

Corpora in LiLa

Linguistic annotation

From Corpora to Lexica

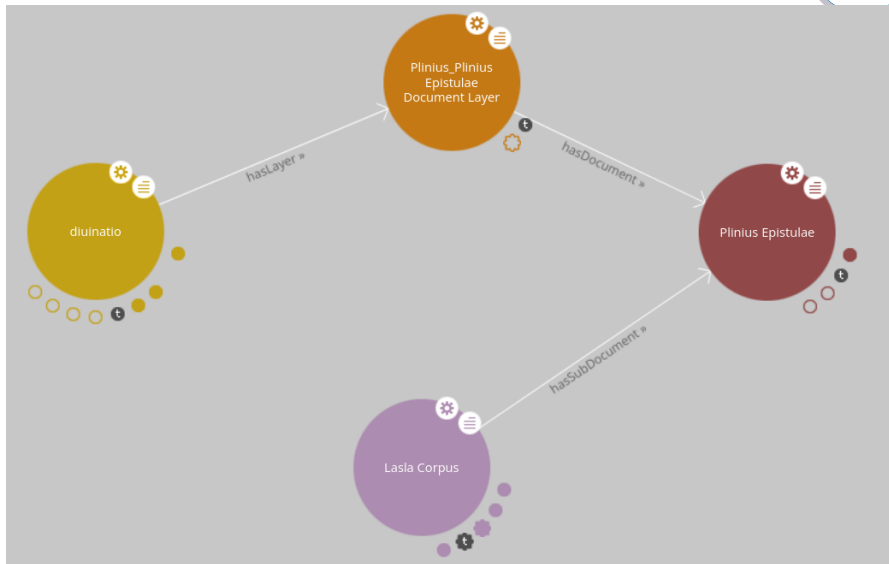
Some queries

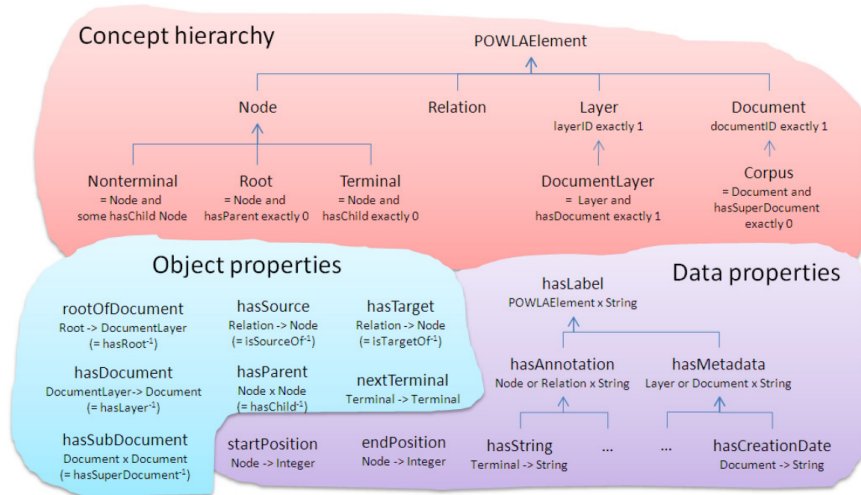
- ▶ Corpus Fibonacci: one chapter from *Liber Abbaci* (lemmatization)
- ▶ LASLA: *Opera Latina*, 131 texts by 20 authors, of classical Latin (lemmatization, Morph. coming soon)
- ▶ ITTB: *Index Thomisticus, Summa Contra Gentiles* (full treebank in 2 format: original and UD)
- ▶ UDante: 5 Latin works of Dante Alighieri (full treebank, UD)

Can we figure out *how many tokens* there are in each of these 4 corpora?

# From Token to Corpora

Via document layers

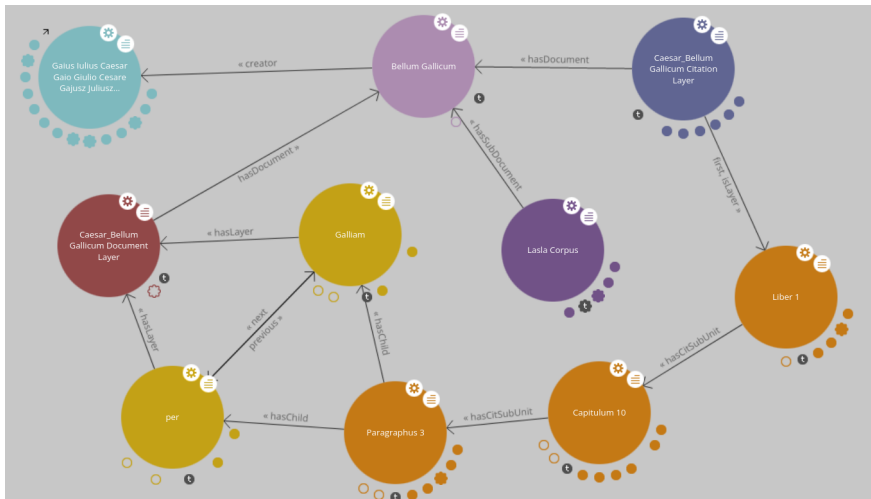




**Fig. 2.** The POWLA ontology (fragment)

# Layers and Metadata

<http://lila-erc.eu/data/id/lemma/109224>





## Introduction

Corpora in LiLa

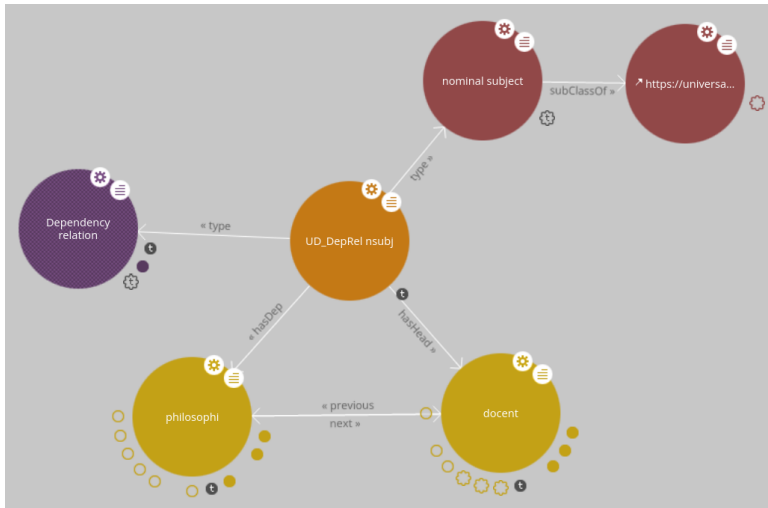
Linguistic annotation

From Corpora to Lexica

Some queries

# Syntactic annotation

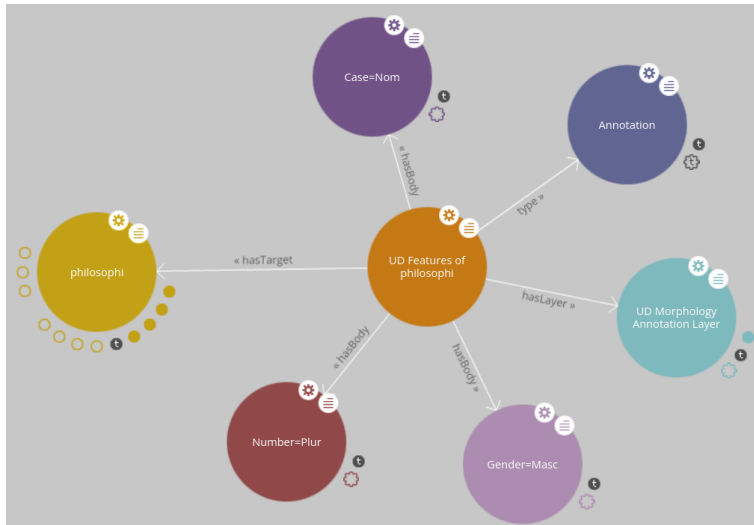
Available for: ITTB, UDante



[http://lila-erc.eu/data/corpora/ITTB/depAnnotation/UD/005.SCG\\*LB4.CP-2++4.N.-8.25-3.27-1W5](http://lila-erc.eu/data/corpora/ITTB/depAnnotation/UD/005.SCG*LB4.CP-2++4.N.-8.25-3.27-1W5)

# Morph features

Available for: ITTB, UDante



[http://lila-erc.eu/data/corpora/ITTB/UD\\_Morphology\\_Annotation/id/token/005.SCG\\*LB4.CP-2++4.N.-8.25-3.27-1W5](http://lila-erc.eu/data/corpora/ITTB/UD_Morphology_Annotation/id/token/005.SCG*LB4.CP-2++4.N.-8.25-3.27-1W5)

## Introduction

Corpora in LiLa

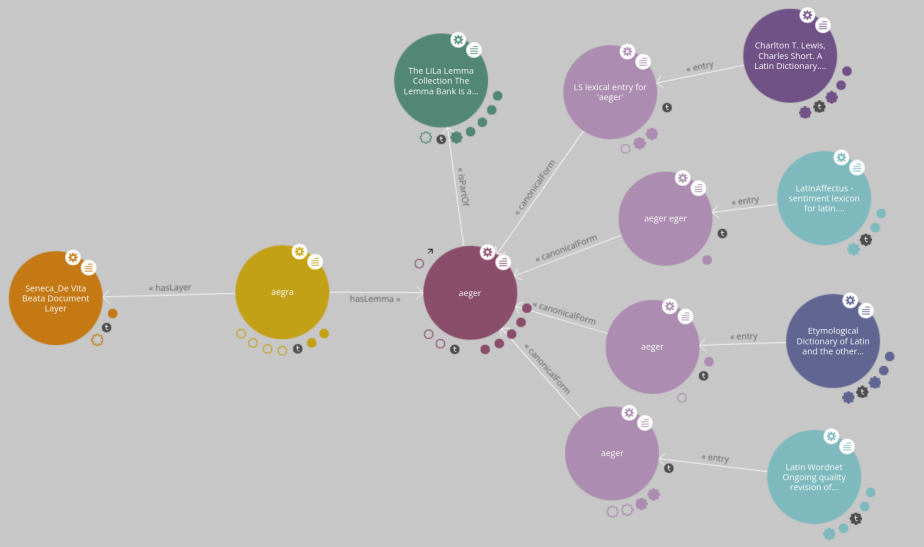
Linguistic annotation

From Corpora to Lexica

Some queries

# Lemmatization as a gateway

*aeger*, 'ill, diseased'



- ▶ Subjects in Thomas and Dante
- ▶ that have a **negative** polarity

*Evil is...*

- ▶ Nominal predicates in Thomas and Dante
- ▶ where the subject is a word with **negative** polarity