



HAL
open science

Zoo-KG: A knowledge graph of zoological data

Arnaud Barbe, Molka Tounsi Dhouib, Catherine Faron, Marco Corneli,
Arnaud Zucker

► **To cite this version:**

Arnaud Barbe, Molka Tounsi Dhouib, Catherine Faron, Marco Corneli, Arnaud Zucker. Zoo-KG: A knowledge graph of zoological data. AIUCD 2023 - Associazione per l'Informatica Umanistica e la Cultura Digitale, Jun 2023, Sienna (Italy), Italy. hal-04235805

HAL Id: hal-04235805

<https://hal.science/hal-04235805v1>

Submitted on 10 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Zoo-KG: A knowledge graph of zoological data

A.Barbe¹, M.Tounsi Dhouib², C.Faron³, M.Corneli⁴, A.Zucker⁵

¹ Université côte d'Azur, CEPAM, CNRS, France, arnaud.barbe@univ-cotedazur.fr

² Université côte d'Azur, INRIA, CNRS, I3S France, dhouib@i3s.unice.fr

³ Université côte d'Azur, INRIA, CNRS, I3S, faron@i3s.unice.fr

⁴ Université côte d'Azur, CEPAM, CNRS, France, marco.corneli@univ-cotedazur.fr

⁵ Université côte d'Azur, CEPAM, CNRS, France, arnaud.zucker@univ-cotedazur.fr

ABSTRACT

In this paper, we present the ZooKG-Pliny knowledge Graph constructed from a manual annotation of Pliny's *Naturalis Historia* using concepts gathered in the thesaurus TheZoo. ZooKG-Pliny is based on a semantic model that formalizes knowledge about the annotations of zoological information in texts. ZooKG-Pliny allows the integration and the interrogation of relevant knowledge in order to support epistemologists, historians and philologists in their analysis of these texts and knowledge transmission through them.

KEYWORDS

History of Zoology, Semantic annotation of Latin texts, RDF knowledge graph, Ancient animals, Linked data and vocabularies.

1. INTRODUCTION

Historians and philologists have to deal every day with an enormous amount of textual resources. Despite the digitization efforts, the proposed tools do not meet the epistemological requirements. They are often limited to lexical and quantitative search on the data. Researchers express a real need for more intelligent tools in order to carry out more elaborate searches that require richer semantic annotation. A typical approach to integrating and exploiting text annotations is to formalize them as a knowledge graph based on linked data and Semantic Web models[1,2,3]. In this context, the objective of our work is to transform into a knowledge graph (ZooKG-Pliny) the manual annotations of *Naturalis Historia* by Pliny the Elder, made by only one linguist researcher and using concepts from TheZoo thesaurus. The ZooKG-Pliny allows the integration and querying of relevant knowledge to provide researchers with richer automated search capabilities to better meet their needs.

3. ZOO-KG: ZOOLOGY KNOWLEDGE GRAPH - PLINY'S NATURALIS HISTORIA

In this work, we considered the books 8 to 11 of Pliny's « *Naturalis Historia* » on ancient zoology. These four books are divided into 911 paragraphs, which are manually annotated by linguists with concepts from the TheZoo thesaurus. In order to automatically exploit these annotations, we lifted the text and the manual annotations into an RDF dataset and exploited it with SPARQL queries. The first step is to automatically generate an RDF dataset describing the semantics of the decomposition of Pliny's work into chapters and paragraphs. The second step consists in integrating the annotations of the text into this RDF dataset. Figure 1 illustrates an example RDF graph representing the fourteen paragraph of chapter 11 *Naturalis Historia*. Thanks to the link between the annotation and the TheZoo thesaurus, we can enrich our knowledge graph by exploiting other existing links in TheZoo. For example, the concept `idc:5066` which denotes a « Tiger » is a sub-concept of the one denoting « Pantherinae ».

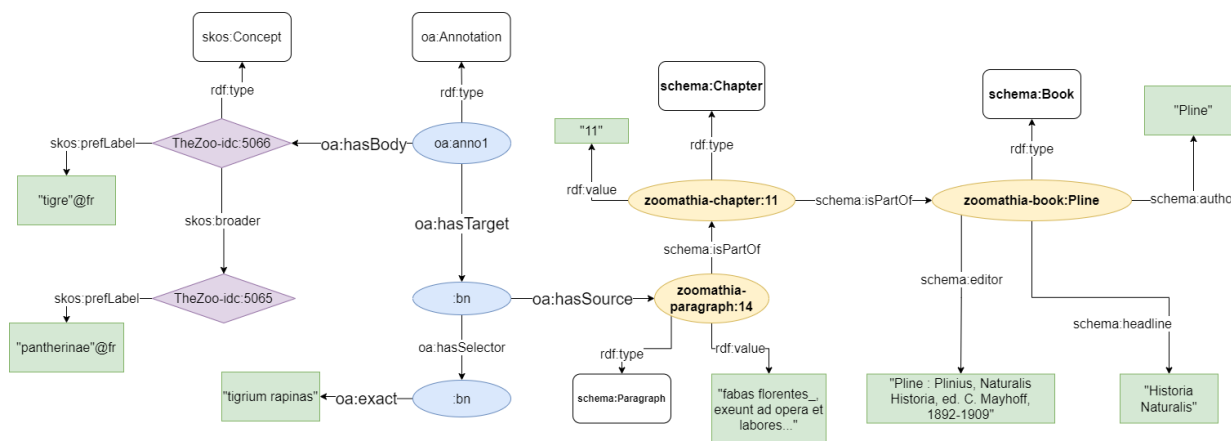


Figure1: Example RDF graph representing the annotation of « *tigrum rapinas* » in paragraph 14 of chapter 11 of Pliny's « *Naturalis Historia* »

At the end of this process, we were able to automatically extract 11073 concepts from the expert's manual annotations, and generated 88184 RDF triples. Table 1 summarizes the number of manual annotations and other key statistics on the resulting knowledge graph.

Number of paragraphs	911
Number of comments	7283
Number of concepts to look for	13241
Number of concepts found	11073
Number of concepts not found	2478
Number of RDF triples generated	88184

Table1: Number of entities per class in our knowledge graph

To illustrate how we can exploit ZooKG-Pliny, we developed a set of SPARQL queries available on the project's Github repository¹. We first collected and discussed a set of competency questions (CQs) formulated by an expert in order to precisely understand their expectations and needs. As an example, Figure 2 presents researchers need to identify in the scientific literature animals that are capable of building a favorable habitat suitable for their needs.

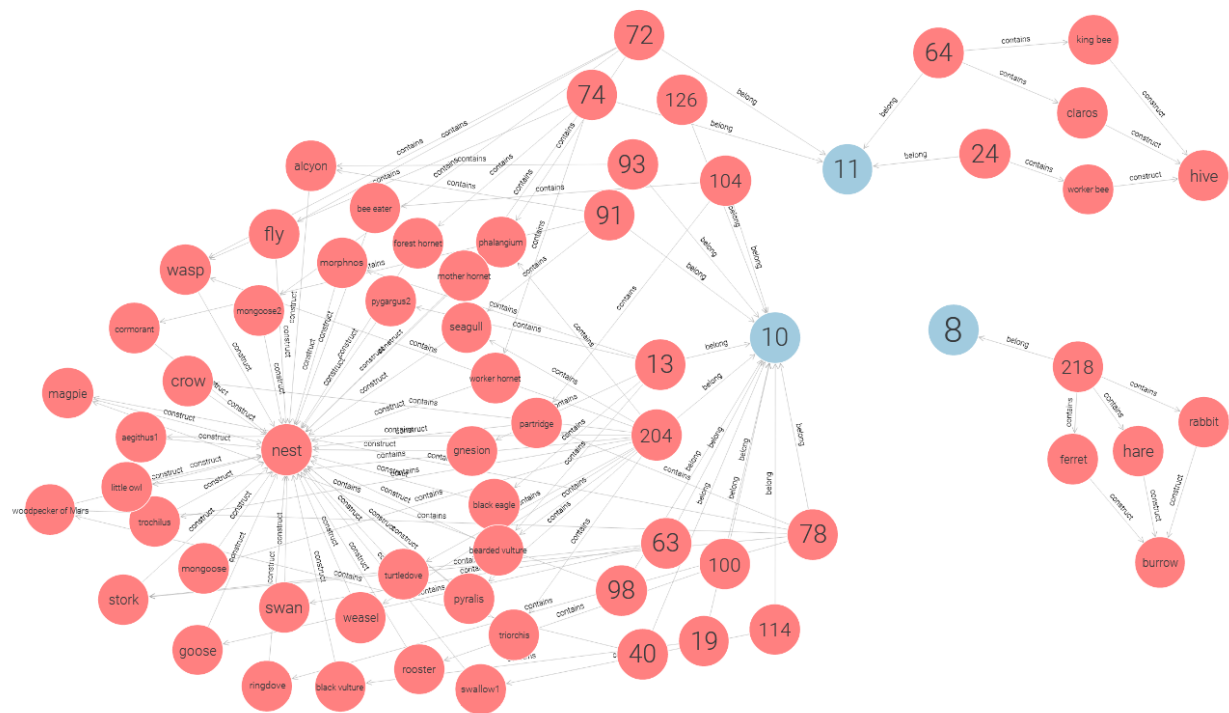


Figure 2: Example of visualization showing the set of animals and their associated constructions cited in Pliny's chapters.

4. CONCLUSION

ZooKG-Pliny is a knowledge graph formalizing manual annotations of four books (8 to 11) of the *Naturalis Historia* by Pliny the Elder. We released a fully automatic pipeline that structures manually extracted information from text and adds a semantic layer for smarter search. As perspectives of this work, we plan to automate or support the manual annotation task that remains tedious for experts and to develop an application that will make it easier for domain experts who are not semantic web specialists to use the constructed knowledge graph via more intelligible or intuitive visualization interfaces.

¹ <https://github.com/Wimmics/zoomathia/tree/main/Pliny>

REFERENCES

- [1] Hyvönen, Eero. "Digital humanities on the Semantic Web: Sampo model and portal series." *Semantic Web Preprint* (2022): 1-16.
- [2] Leyra, Irene Pajón, Arnaud Zucker, and Catherine Faron Zucker. "TheZoo: un thesaurus de zoologie ancienne et médiévale pour l'annotation de sources de données hétérogènes." *Archivum Latinitatis Medii Aevi* 73 (2015): 321-342.
- [3] Tounsi, Molka, Catherine Faron Zucker, Arnaud Zucker, Serena Villata, and Elena Cabrio. "Studying the history of pre-modern zoology with linked data and vocabularies." In *The First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference*. 2015.