Construction d'un graphe de connaissances à partir des annotations manuelles de textes de zoologie antique

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

Introduction

Objectifs

- Collaboration entre Wimmics et CEPAM
- Transmission des savoirs zoologiques
- Définition d'un modèle pour représenter un corpus d'annotations manuelles
  - Extraction des entités à partir des annotations
  - Liage avec des concepts de vocabulaire de domaine
  - Transformation automatique en RDF
  - Implémentation des questions de compétences

Questions de recherche

1. Quels types de connaissances devons-nous représenter afin d'aider les chercheurs dans leur travail d'analyse de la transmission de savoirs zoologiques à travers les textes ?

2. Quels vocabulaires existants pouvons-nous réutiliser pour représenter ces documents ?

3. Quelle approche pouvons-nous définir pour réutiliser les annotations manuelles faites par les linguistes et les rendre exploitables automatiquement?

Description du corpus de textes annotés

Histoire naturelle chapitres 8-11

obedientia, officiorum quae didicere memoria, amoris et

gloriae voluptas, immo vero, quae etiam in homine rara,

probitas, prudentia, aequitas, religio quoque siderum solis que

ac lunae veneratio.

[[2]] Auctores sunt in Mauretaniae saltibus ad quendam

amnem, cui nomen est Amilo, nitescente luna nova greges

Paragraphe

Annotations = commentaire word

Auteur	memory
Auteur	pleasure
Auteur	comparison with human
Auteur	sense of justice

Ⓐ Auteur
religion : astral cult : moon cult , sun cult

🔁 Répondre   💬 Résoudre

Auteur	places : Libya : region :
Auteur	habitat : terrestrial habitat :
Auteur	places : Libya : Mauretania :
Auteur	intellectual authority : undefined
Auteur	time : moon phase : new moon
Auteur	gregarious

Motifs d'annotation manuelle

Motifs d'annotation	Exemple
"concept"	elefant
"concept1; concept2; ..."	old age ; social hierarchy
"concept1 : concept2 : ..."	habitat: terrestrial:forest
"concept1 : concept2, concept3, ..."	historical character :grammarian , Aristophanes of Byzance; anecdote
"collection : concept"	Places : Italia
Combinaisons de tous ces motifs	historical character ; intellectual authority : philosopher : Antipater

Questions de compétences

- Quels sont les animaux qui construisent un habitat ?

- Quels sont les passages du texte qui permettent d'identifier les interactions possibles entre l'homme et l'animal (la prédation, l'amitié et l'animosité)?

- Quels sont les remèdes (thérapeutiques) dont un ingrédient est une partie d'animal ?

Vocabulaires existants

[[TITRE]] *Historia naturalis,* LIBER VIII

[[AUTEUR]] *Pline*

[[EDITION]] Pline : Plinius, Naturalis Historia, ed. C.

Mayhoff, 1892-1909

 [[1]] Ad reliqua transeamus animalia et primum terrestria.

Maximum est elephans proximum que humanis sensibus, quippe intellectus illis sermonis patrii et imperiorum obedientia, officiorum quae didicere memoria, amoris et gloriae voluptas, immo vero, quae etiam in homine rara, probitas, prudentia, aequitas, religio quoque siderum solis que

schema.org

Web Annotation Vocabulary

Auteur
10:50 Aujourd'hui
size : relative

À partir d'un document importé.

Auteur
10:50 Aujourd'hui
elefant

TheZoo

Thesaurus TheZoo

- Représenter et structurer hiérarchiquement tous les concepts de la zoologie antique et médiévale à partir de :
  - ressources textuelles
  - ressources iconographiques
  - ressources archéologiques

- 6019 concepts

- 11 niveaux hiérarchiques

- 14 collections

Modèle proposé
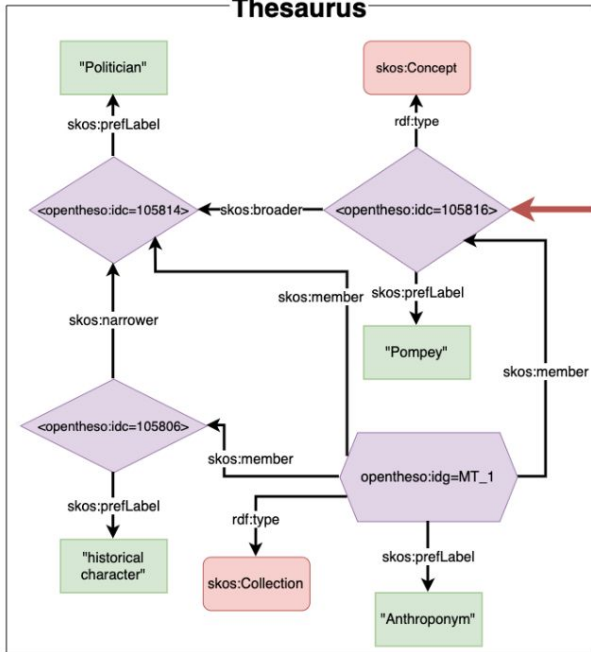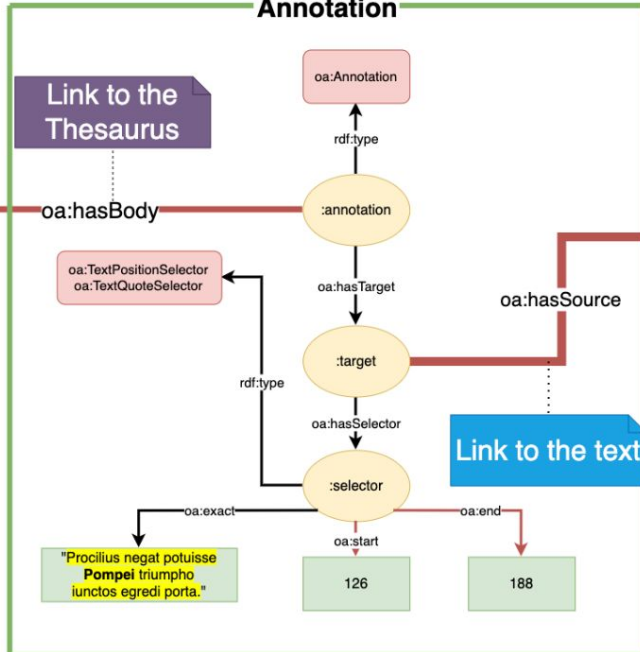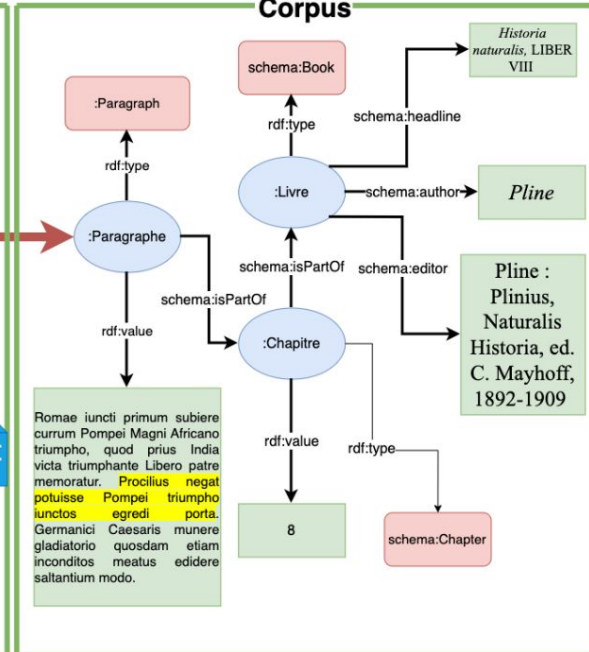
Processus de construction du graphe de connaissance
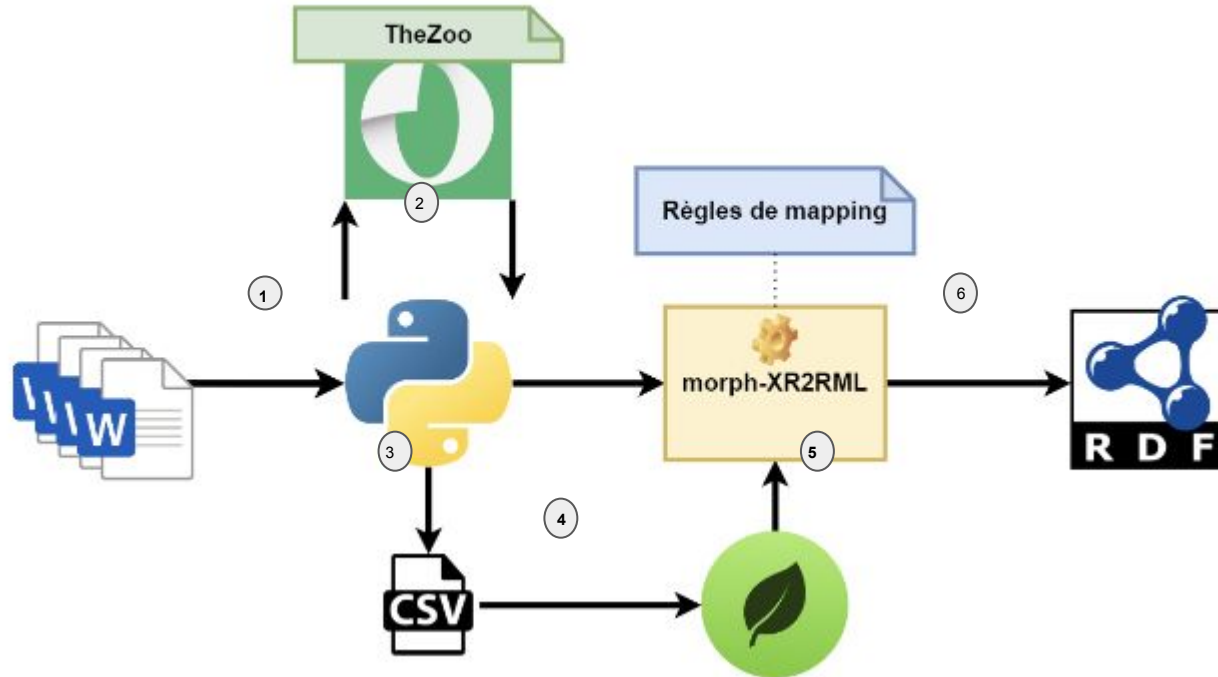
Description du graphe de connaissances produit

Nbre de paragraphes	911
Nbre de commentaires	7283
Nbre d'entités reconnues	13241
Nbre d'entités liées	11590
Nbre d'entités non liées	2632
Nbre de triplets RDF générés	88184

Evaluation du graphe produit

Evaluation du processus d'extraction de connaissances

- Précision = 1

- Rappel= 0.814

Evaluation du graphe produit

Implémentation en SPARQL des questions de compétence

Quels sont les animaux qui construisent un habitat?

```sparql
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?paragraph ?name_animal  ?name_construction
WHERE {?annotation1 a oa:Annotation; oa:hasBody ?animal;
             oa:hasTarget [oa:hasSource ?paragraph; oa:hasSelector [oa:exact?mention_animal]].
?annotation2 oa:hasBody ?construction; oa:hasTarget [oa:hasSource ?paragraph;
             oa:hasSelector [oa:exact ?mention_construction]].
?animal     a skos:Concept; skos:prefLabel ?name_animal.
<https://opentheso.huma-num.fr/?idg=MT_10&idt=th310> skos:member ?animal.
?construction skos:prefLabel ?name_construction;
             skos:broader+ <https://opentheso.huma-num.fr/?idc=105466&idt=th310>.
FILTER (lang(?name_animal) = "en"). FILTER (lang(?name_construction) = "en")
}ORDER BY ?paragraph
```

Paragraph	Text	Animal Concept	Mention Animal	Construction concept	Mention construction
http://www.zoomathia.co 10-annoteIPL/10/	Plurimum volant quae apodes, quia careant usu pedum, ab aliis cypseli appellantur, hirundinum specie. nidificant in scopulis. hae sunt quae toto mari cernuntur, nec umquam tam longo naves tamque continuo cursu recedunt a terra, ut non circumvolitent eas apodes. cetera genera residunt et insistunt; his quies nisi in nido nulla: aut pendent aut iacent.	swallow1	hirundinum specie	nest	in nido

Lors d'un clic d'une ligne

Chapitre 10

[114] Plurimum volant quae apodes, quia careant usu pedum, ab aliis cypseli appellantur, hirundinum specie, nidificant in scopulis. hae sunt quae toto mari cernuntur, nec umquam tam longo naves tamque continuo cursu recedunt a terra, ut non circumvolitent eas apodes. cetera genera residunt et insistunt; his quies nisi in nido nulla: aut pendent aut iacent.

swallow1

nest

14

Conclusion

- Construction d'un graphe de connaissances à partir des annotations manuelles de l'Histoire Naturelle de Pline l'Ancien

- Capture du contexte d'apparition des différentes annotations et structuration avec Open Annotation

- Liage des annotations manuelles avec TheZoo

- Interrogation du graphe en exploitant les contextes d'apparition et les liens entre les concepts du TheZoo

- Des problèmes de modélisation soulevés au niveau de TheZoo

Perspectives

- Automatisation de la tâche d'annotation des textes

- Développement d'interfaces de visualisation plus intelligibles et intuitives pour les experts

Merci pour votre attention!

Extraction et désambiguïsation d'entités nommées

Requête SPARQL de recherche de concepts dans TheZoo à partir de labels

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT ?x WHERE {
{?x a ?type; skos:prefLabel ?label.
 ?y skos:prefLabel ?collection; skos:member ?x.
 FILTER(lang(?label) = "en").
 FILTER("{label}" in (ucase(str(?label)),lcase(str(?label)), str(?label))).
 FILTER("{parent}" in (ucase(str(?collection)), lcase(str(?collection)), str(?collection))).
} UNION {?x a ?type; skos:prefLabel ?label;skos:broader+ ?y.
 ?y skos:prefLabel ?concept;
FILTER(lang(?label) = "en").
FILTER("{label}" in (ucase(str(?label)),lcase(str(?label)), str(?label))).
FILTER("{parent}" in (ucase(str(?concept)),lcase(str(?concept)), str(?concept))).
}}
```