

A unified Knowledge Graph based Approach to Exploring Scientific Text Corpora

Catherine Faron



séminaire HislNum, Académie 5, 25/09/2024

Wheat Genomics Scientific Literature Knowledge Graph



Data to Knowledge in Agronomy and Biodiversity
2019-2024
ANR grant ANR-18-CE23-0017
<http://www.d2kab.org>

Researchers need help to make sense of scientific literature

The screenshot shows the PubMed search interface with the search term "wheat genomics" entered. The results page displays 15,620 results, which is highlighted with a red oval. The results are sorted by "Best match". On the left, there are filters for "RESULTS BY YEAR" (a histogram from 1964 to 2024), "TEXT AVAILABILITY" (Abstract, Free full text, Full text), and "ARTICLE ATTRIBUTE" (Associated data). Below the search bar, there are buttons for "Save", "Email", and "Send to". The main content area shows three article snippets:

- The genetic basis of wing spots in Pieris canidia butterflies.** Wee JLQ, et al. *BMC Genomics*. 2023. PMID: 37016295 [Free PMC article](#).
- Quantitative RNA-Seq analysis in non-model species: assessing transcriptomes as a scaffold and the utility of evolutionary divergent genome assemblies.** Hornett EA, et al. *BMC Genomics*. 2012. PMID: 22853326 [Free PMC article](#).
- Short read Illumina data for the de novo assembly of a non-model snail transcriptome (Radix balthica, Basommatophora, Pulmonata), and a comparison of assembler performance.** Feldmeyer B, et al. *BMC Genomics*. 2011. PMID: 21679424 [Free PMC article](#).

On the right side of the results page, there is a "Tracking" section with a snippet from "Theor Appl Genet." and a "Complex relationship between DNA methylation and gene expression due to Lr28 in wheat-leaf rust" article by Gautam Saripalli et al. The right sidebar includes links for "SpringerLink FULL TEXT ARTICLE", "Cite", "Collections", "SHARE" (with icons for X, f, and g), and "PAGE NAVIGATION" (Title & authors, Abstract).



|

Advanced

Search

User Guide

> Mol Biol Rep. 2020 Feb;47(2):1339-1360. doi: 10.1007/s11033-019-05236-1. Epub 2019 Dec 23.

Complex relationship between DNA methylation and gene expression due to Lr28 in wheat-leaf rust pathosystem

Gautam Saripalli ¹, Chanchal Sharma ^{1 2}, Tinku Gautam ¹, Kalpana Singh ³, Neelu Jain ⁴,
Pramod Prasad ⁵, J K Roy ⁶, J B Sharma ⁴, P K Sharma ¹, K V Prabhu ^{4 7}, H S Balyan ^{1 3},
P K Gupta ⁸

Affiliations + expand

PMID: 31873872 DOI: 1

Abstract

Differential DNA methylation in resistant (R) near isogenic amplified polymorphism had a large number of hypermethylated genes during the passage

FULL TEXT LINKS



ACTIONS

“ Cite

Collections

Gene

Differential DNA methylation due to Lr28 was examined in susceptible (S) wheat cv. HD2329 and its resistant (R) near isogenic line (NIL) (HD2329+Lr28) using two approaches: methylation sensitive amplified polymorphism (MSAP) and methylated DNA immunoprecipitation (MeDIP) (...). The results of the present study improved our understanding of the epigenetic control of leaf rust resistance in wheat.

Trait

Taxon

When R NIL was compared with S cultivar, there were many hypermethylated and fewer hypomethylated genes in R NIL relative to S cultivar, suggesting that many genes that are active in S cultivar are silenced in R NIL both at 0 hpi and at 06 hpi. Level of methylation was generally abundant

Similar articles

> Mol Biol Rep. 2020 Feb;47(2):1339-1360. doi: 10.1007/s11033-019-05236-1. Epub 2019 Dec 23.

Complex relationship between DNA methylation and gene expression due to Lr28 in wheat-leaf rust pathosystem

Gautam Saripalli ¹, Chanchal Sharma ^{1 2}, Tinku Gautam ¹, Kalpana Singh ³, Neelu Jain ⁴,
Pramod Prasad ⁵, J K Roy ⁶, J B Sharma ⁴, P K Sharma ¹, K V Prabhu ^{4 7}, H S Balyan ^{1 3},
P K Gupta ⁸

Affiliations + expand

PMID: 31873872 DOI: 1

Abstract

Differential DNA methylation in resistant (R) near isogenic amplified polymorphism had a large number of hypermethylated genes during the passage

When R NIL was compared with S cultivar, there were many hypermethylated and fewer hypomethylated genes in R NIL relative to S cultivar, suggesting that many genes that are active in S cultivars are silenced in R NIL both at 0 hpi and at 06 hpi. Level of methylation was generally abundant

FULL TEXT LINKS



ACTIONS

“ Cite

_collections

Gene

Differential DNA methylation due to Lr28 was examined in susceptible (S) wheat cv. HD2329 and its resistant (R) near isogenic line (NIL) (HD2329+Lr28) using two approaches: methylation sensitive amplified polymorphism (MSAP) and methylated DNA immunoprecipitation (MeDIP) (...). The results of the present study improved our understanding of the epigenetic control of leaf rust resistance in wheat.

Trait

Wheat Trait and Phenotype Ontology (WTO)

Similar articles

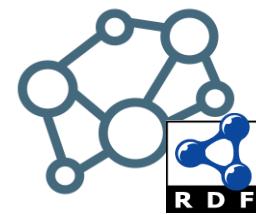
Taxon



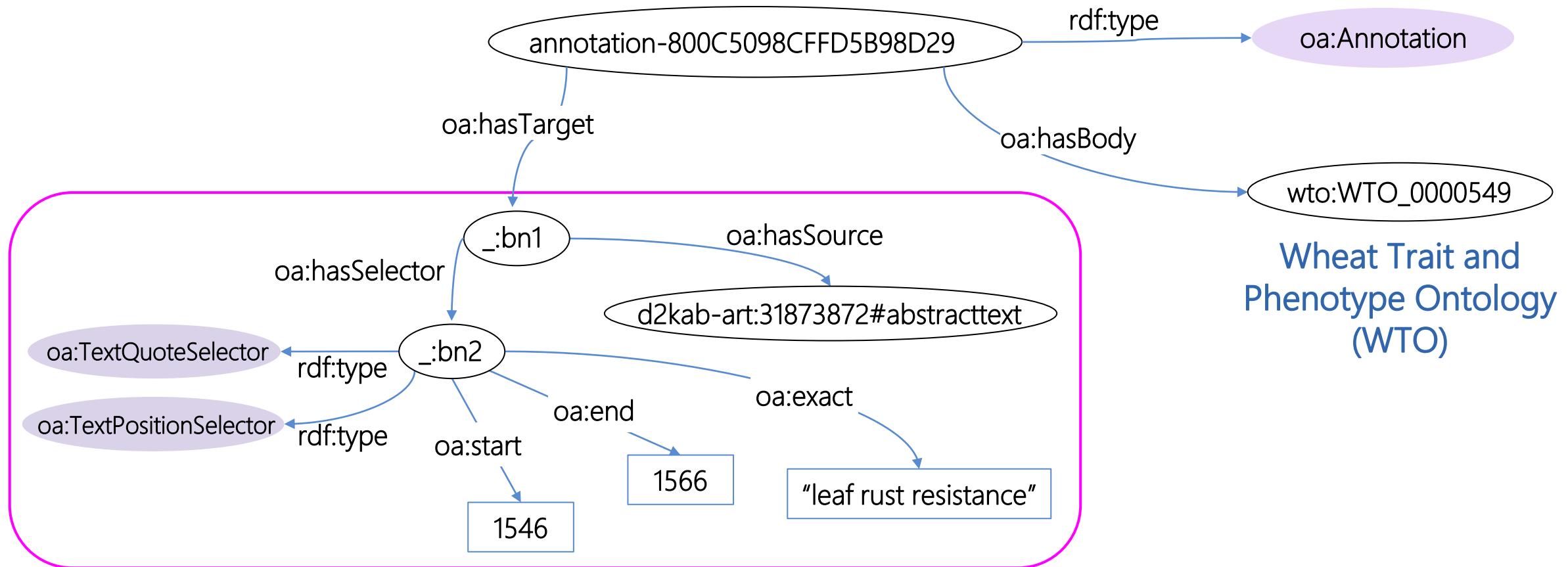
Taxonomy

Extracting & Linking Named Entities

Modeling mentions of terms as annotations



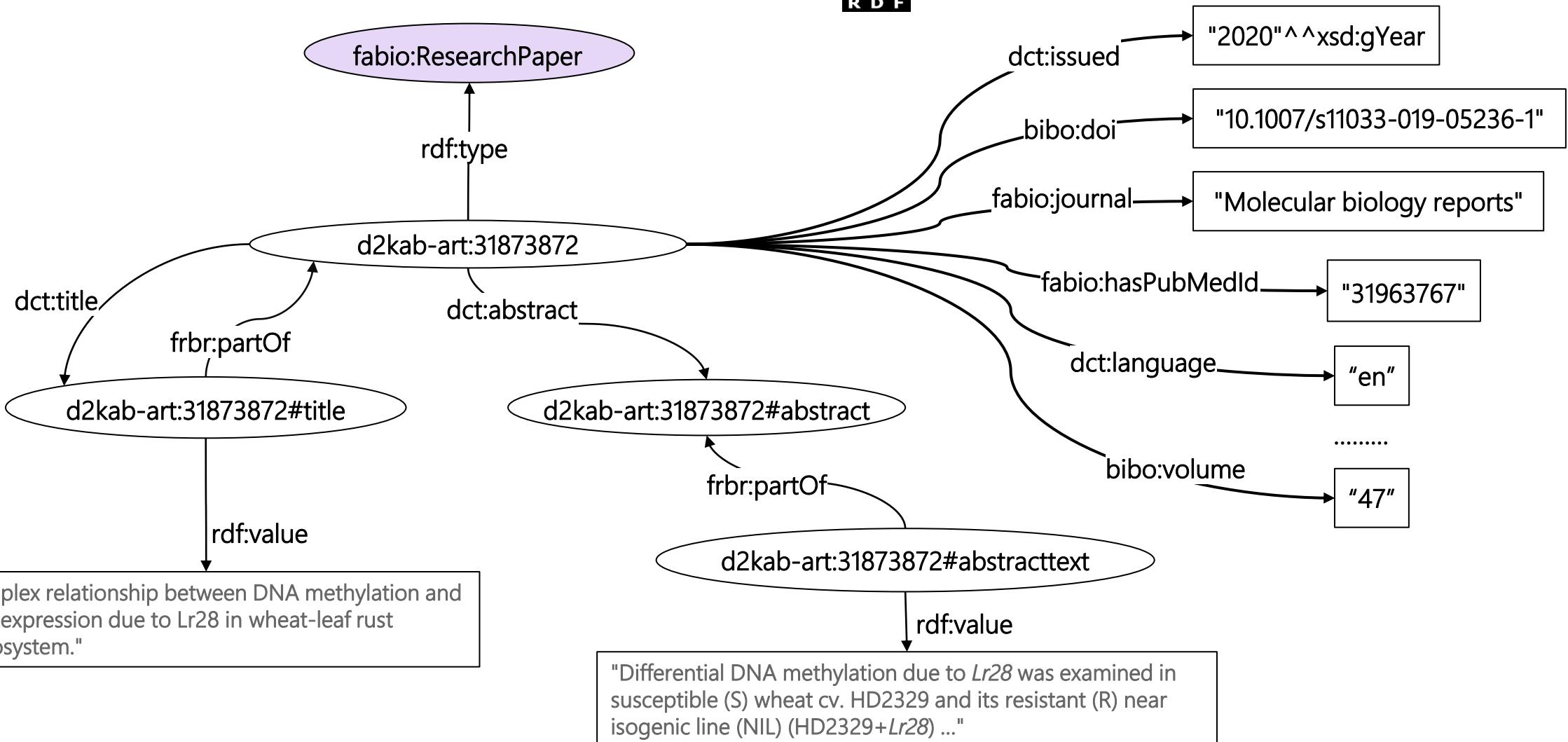
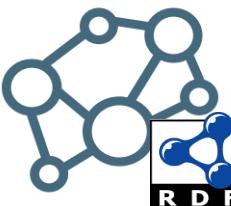
Web Annotations Ontology 



Wheat Trait and Phenotype Ontology (WTO)

Tool: Morph-xR2RML - <https://github.com/frmichel/morph-xr2rml/>

Modeling articles metadata



Tool: SPARQL micro-service - <https://github.com/frmichel/sparql-micro-service/>

WheatGenomicsSLKG statistics

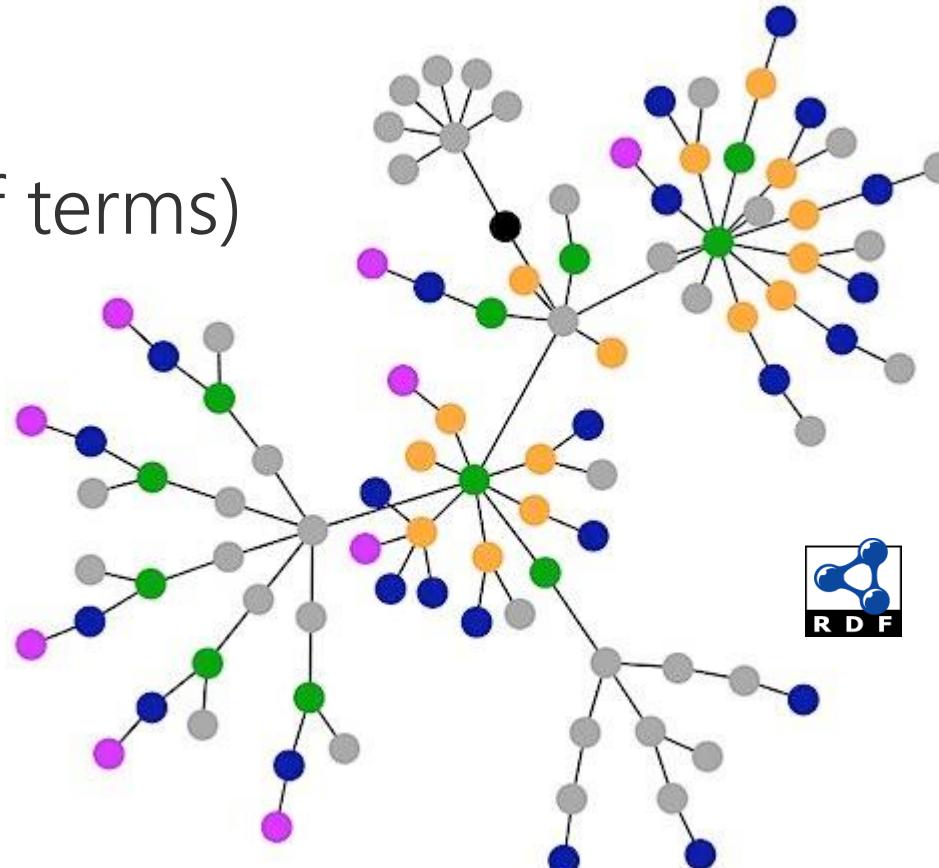
1,191,867 triples

8,496 scientific articles

88,880 annotations (mentions of terms)

4,319 unique entities:

- 2463 taxa
- 98 traits
- 77 varieties
- 1160 genes
- 521 genetic markers





Search articles by named entities

Enter text and select a suggestion

[Search](#)[rust resistance](#) x[Lr34](#) x

Results for only the selected named entities

Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in wheat.

QTL for spot blotch resistance in bread wheat line Saar co-locate to the biotrophic disease resistance genes Lr34 and Lr46.

Theoretical and applied genetics. Theoretische und angewandte Genetik.

The Lr34 adult plant rust resistance gene provides seedling resistance in durum wheat.

Postulation of rust resistance genes in Nordic spring wheat genotypes and identification of new resistance genes.

Bariana H, Lillemo M [...]. 2016. *Journal of applied genetics*.

Gene expression patterns in near isogenic lines for wheat rust resistance gene Lr34/yr.

Results for the selected named entities or their sub-entities

[resistance to Stem rust](#) [Lr34](#)

Leaf rust resistance gene Lr34 associated with nonsuppression of stem rust resistance.

[resistance to Stripe rust](#) [Lr34](#)

New slow-rusting leaf rust and stripe rust resistance genes Lr67 and Yr46 in wheat and barley. [...]. 2011. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*.



Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in wheat.

Kwiatek M, Nawraćała J, Skowrońska R, Tomkowiak A. 2019. Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in wheat.. *Journal of applied genetics*.

Language: eng

[Read the article](#)

Abstract

Leaf rust caused by *Puccinia triticina* belongs to one of the most dangerous fungal diseases of wheat (*Triticum aestivum L.*) and is the cause of large yield losses every year. Here we report a multiplex polymerase chain reaction (PCR) assay, which was developed for detection of two important wheat slow rust resistance genes Lr34 and Lr46, using two molecular markers: csLV34 and XwmC44, respectively. The presence of genes was analyzed in one winter wheat variety TX89D6435 and five spring wheat varieties: Pavon F76, Parula 'S', Rayon 89, Kern, Mochis 88. Both Lr34 and Lr46 genes were identified in variety TX89D6435, gene Lr34 was also identified in Parula 'S' and Kern varieties, and gene L46 occurs in Pavon F76 and Mochis 88 variety. None of the resistance genes tested was detected in the Rayon 89 variety. The use of the multiplex PCR method allowed to shorten the analysis time, reduce costs of analyses, and reduce the workload.

[Hide named entities](#)

Linked Data Explorer

Query name:

[Wimmics team](#) Inria, I3S, UCA, CNRS.

Initial Query

SPARQL Endpoint * <http://d2kab.i3s.unice.fr/sparql>

Query * Query 2. Co-occurrence of a trait

Concept 1 resistance to rust

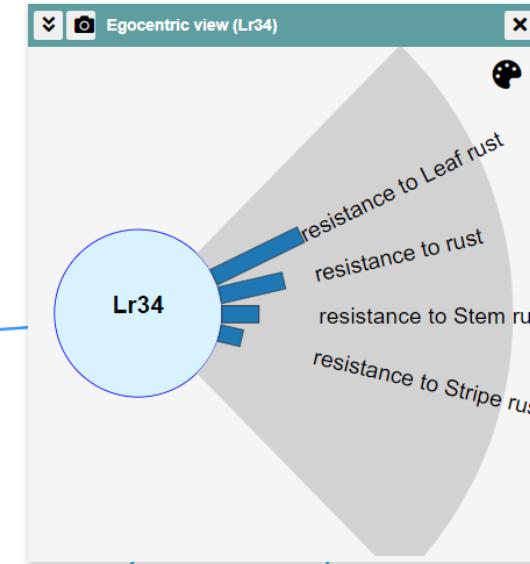
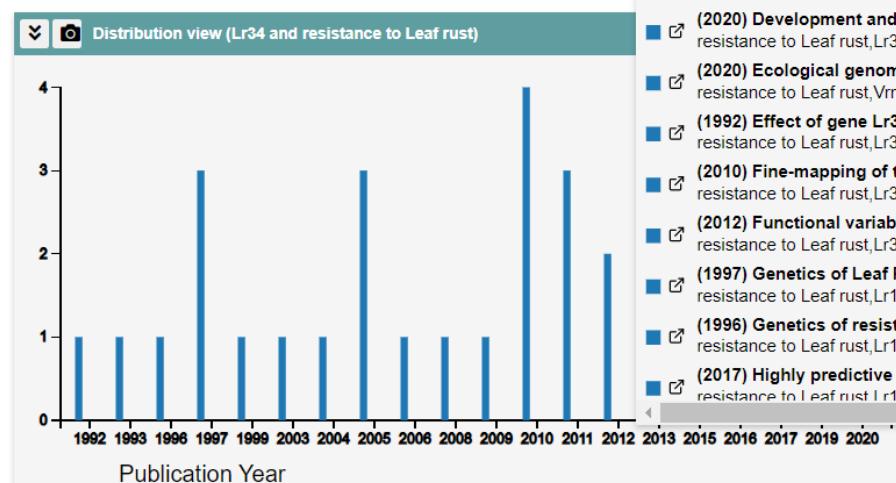
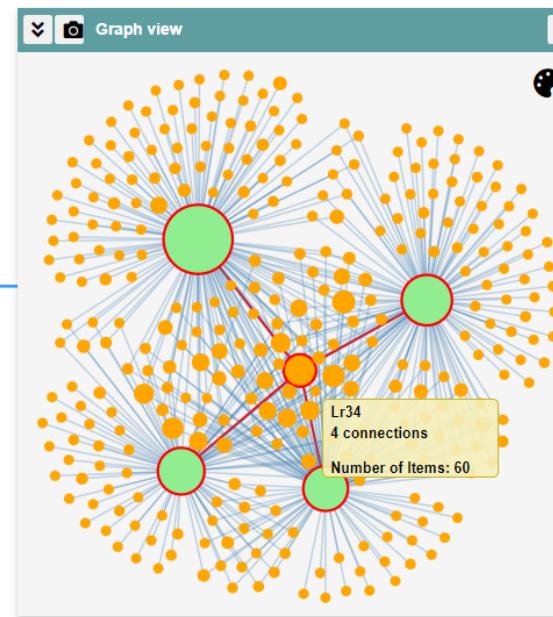
Use Stylesheet

History

- Initial Query
- Graph view
- Egocentric view (Lr34)
- Distribution view (Lr34 and resistance to Leaf rust)
- Listing view (Lr34 and resistance to Leaf rust)

Clear cache Run

Save



- Listing view (Lr34 and resistance to Leaf rust)**
- (2011) A multiple resistance locus on chromosome arm 3BS in wheat resistant to Leaf rust, Sr2, Lr27, resistance to Stem rust, Lr34
 - (2020) Development and deployment of KASP markers for multiple allele resistance to Leaf rust, Lr34
 - (2020) Ecological genomics of Chinese wheat improvement: implications for resistance to Leaf rust, Vrn-B1, Lr34, Ppd-D1, Glu-B1
 - (1992) Effect of gene Lr34 in the enhancement of resistance to leaf rust resistance to Leaf rust, Lr34, Lr18
 - (2010) Fine-mapping of the leaf rust Lr34 locus in *Triticum aestivum* (L.) resistance to Leaf rust, Lr34
 - (2012) Functional variability of the Lr34 durable resistance gene in transgenic wheat resistance to Leaf rust, Lr34
 - (1997) Genetics of Leaf Rust Resistance in Canadian Spring Wheats AC resistance to Leaf rust, Lr16, Lr34, Lr12, Lr13, LrTb
 - (1996) Genetics of resistance to wheat leaf rust. resistance to Leaf rust, Lr13, Lr34
 - (2017) Highly predictive SNP markers for efficient selection of the wheat resistance to Leaf rust Lr16 Lr34

WheatGenomicsSLKG

Dataset (RDF dump)	10.5281/zenodo.10410742
SPARQL endpoint	http://d2kab.i3s.unice.fr/sparql
Documentation	https://github.com/Wimmics/WheatGenomicsSLKG
URIs namespace	http://ns.inria.fr/d2kab/
Web interfaces	http://d2kab.i3s.unice.fr/wheatgenomics/search https://dataviz.i3s.unice.fr/mgexplorer/dashboard with SPARQL endpoint http://d2kab.i3s.unice.fr/sparql
Publications	<p>Yacoubi Ayadi N., Faron C., Michel F., Bossy R., Barbe A. Construction d'un graphe de connaissances à partir des annotations d'articles scientifiques et de leur contenu en sciences de la vie. IC 2022 - PFIA 2022. hal-03889968</p> <p>Yacoubi Ayadi N., Bernard S., Bossy R., Courtin M., Happi Happi B.G., Larmande P., Michel F., Nédellec C., Roussey C., Faron C. A unified approach to publish semantic annotations of agricultural documents as knowledge graphs. Smart Agricultural Technology , vol 8, 2024</p>



Semantic Indexing of a Scientific Archive in Agriculture

ISSA2



Projet Lauréat

CollEx-Persée

<https://issa.cirad.fr/>

Slides from Franck Michel

Researchers need help to explore the Cirad open publication archive

Agritrop, the Cirad open archive

+120 000 bibliographical references including +42 000 in open access; +33 000 articles including +16 in open accessss

A manual indexation by specialized documentalists until the end of 2023

A reference thesaurus for indexing in Agritrop : Agrovoc (thematic and geographic descriptors)



Food and Agriculture Organization of the United Nations



Host plants associated with *Diatraea tabernella* Dyar (Lepidoptera: Crambidae) in sugarcane in Panama

Atencio Valdespino Randy, Goebel François-Régis, Murillo Vielka. 2018. Host plants associated with *Diatraea tabernella* Dyar (Lepidoptera: Crambidae) in sugarcane in Panama. *International Sugar Journal*, **120** (1438) : 786-791.
<https://internationalsugarjournal.com/paper/host-plants-associated-with-diatraea-tabernella-dyar-lepidoptera-crambidae-in-sugarcane-in-panama/>

Article de revue ; Article de recherche



Version publiée - Anglais

Accès réservé aux agents Cirad

Utilisation soumise à autorisation de l'auteur ou du Cirad.

ISJ_Oct18_Goebel_Atencio.pdf

[Télécharger \(273kB\)](#) | [Demander une copie](#)

Résumé : Within the genus *Diatraea*, studies of alternate host plants were mainly conducted on *Diatraea saccharalis* (Fab.). Such information doesn't exist for *Diatraea tabernella* Dyar. Therefore, the objective of this study was to determine the alternative host plants of *D. tabernella* in sugarcane plantations in Panama. From January 2016 to February 2017, a general inventory of alternate host plants was conducted in the sugarcane field and then plants were sampled among the most frequent 9 species found (Poaceae (8) and Cyperaceae (1)) in four areas the sugarcane field (near water source (Z1), within field

(Z2), near mangrove (Z3) and field edge (Z4)). The species with the highest % borer infestation during the sugarcane growth period were *Sorghum halepense* (L.) Pers. (7.2%), *Echinochloa colonum* (L.) Link (6.4%) and *Eleusine indica* (L.) Gaertn. (5.8%). During the sugarcane harvest the species with the highest percentage of infested stems were *Eleusine indica* (L.) Gaertn. (21.6%), *Dactyloctenium aegyptium* (L.) Beauv. (17.1%) and *Cenchrus echinatus* (L.) (9.2%). On these host plants, species control measures may be applied to reduce stemborer populations.

Mots-clés Agrovoc : Canne à sucre, Ravageur des plantes, Lutte antiravageur, Désherbage, Croissance

Mots-clés géographiques Agrovoc : Panama

Mots-clés complémentaires : *Diatraea tabernella*

Mots-clés libres : HOST PLANTS, STALK BORERS, SUGARCANE, Integrated Pest Management, Weed management

Classification Agris : H10 - Ravageurs des plantes

H60 - Mauvaises herbes et désherbage

F62 - Physiologie végétale : croissance et développement

Auteurs et affiliations

- Atencio Valdespino Randy, CIRAD-PERSYST-UPR AIDA (FRA) - auteur correspondant
- Goebel François-Régis, CIRAD-PERSYST-UPR AIDA (FRA)
- Murillo Vielka, Universidad de Panama (PAN)

Source : Cirad-Agritrop (<https://agritrop.cirad.fr/591124/>)

Indexation manuelle

Concept search

- Descripteurs ou Entités nommées
- Multi-langue, synonymes

The screenshot shows the ISSA Agritrop search interface. At the top left is the ISSA logo with a stylized magnifying glass icon. A text block states: "This interface developed by the ISSA project demonstrates the enriched visualization of Agritrop documents." To the right is the Agritrop logo in pink, followed by "Open Repository of CIRAD publications" and a "Document Search" link. Below the header is a search form with two radio buttons: "Agrovoc descriptors" (selected) and "Wikidata named entities". A search input field contains "Enter text and select a suggestion", a "Search" button, and a recent search term "climate change" with a close button. The main content area is titled "Search documents" and displays search results for "climate change". It shows "1929 result(s.)" with a page navigation bar from 1 to 242. The results list four items, each with a document icon and a download icon:

- La transition agro-écologique des agricultures du Sud. Côte, François-Xavier (ed.), Perret, Sylvain (ed.), Poirier-Magona, Emmanuelle (ed.) [...]. 2018.
- Climate, cattle rearing systems and African animal trypanosomosis risk in Burkina Faso. Akoudjin, Massouroudini, Belem, Adrien Marie Gaston, Bengaly, Zakaria [...]. 2012. *PloS One*.
- Perception paysanne des perturbations pluviométriques et stratégies d'adaptation dans les systèmes de culture à sorgho repiqué en zone soudano-sahélienne du Cameroun. Fokou Yemata, Oberline, Madi, Ali, Oumarou, Yakouba [...]. 2017. *Afrique Science*.
- Agroforestry rubber networks and farmers groups in Phatthalung area in Southern Thailand. A potential for an innovation platform?. Chambon, Bénédicte, Michel, Isabelle, Penot, Eric [...]. 2022. *Forest and Society*.

Concept search

- Descripteurs ou Entités nommées
- Multi-langue, synonymes
- Hiérarchie (termes plus spécifiques)

AGROVOC Multilingual Thesaurus

Content language English ▾ Search

Alphabetical Hierarchy Groups

phenomena > natural phenomena > climate change

PREFERRED TERM **climate change**

DEFINITION **climate change** (en) As alterações climáticas são mudanças do clima que ocorrem a nível global. Segundo o "Painel Intergovernamental sobre Mudanças Climáticas" (IPCC, 2001) as alterações climáticas

BROADER CONCEPT natural phenomena (en)

NARROWER CONCEPTS anthropogenic climate change (en) global warming (en)

RELATED CONCEPTS climate change adaptation (en) climate change impacts (en)

air-water exchanges atmospheric circulation atmospheric forcing atmospheric formations atmospheric optical phenomena carbon sinks climate anthropogenic climate change **global warming** (en) climate variability cycling darkness deglaciation ecological succession ecosystem disturbance El Niño environmental degradation environmental impact fire causes forest fragmentation geological processes greenhouse effect

This interface developed by the ISSA project demonstrates the enriched visualization of Agritrop documents.

Agritrop Open Repository of CIRAD publications Document Search

Search documents

Agrovoc descriptors Wikidata named entities

Enter text and select a suggestion Search

climate change ×

Results matching only the selected descriptors

1929 result(s.)

La transition agro-écologique des agricultures du Sud. Côte, François-Xavier (ed.), Perret, Sylvain (ed.), Poirier-Magona, Emmanuelle (ed.) [...], 2018

Results matching the selected descriptors or any more specific descriptors

4 result(s.)

global warming The interactive effect of temperature and fertilizer types determines the dominant microbes in nitrous oxide emissions and the dicyandiamide efficacy in a vegetable soil. Di, Hongjie, Herrmann, Laetitia, Lesueur, Didier [...]. 2024. *Soil Ecology Letters*.

global warming Lowering N2O emissions from soils using eucalypt biochar: the importance of redox reactions. Cowie, Annette, Donne, Scott, Husson, Olivier [...]. 2015. *Scientific Reports*.

global warming Unravelling life cycle impacts of coffee: Why do results differ so much among studies?. Acosta Alba, Ivonne, Azapagic, A., Boissy, Joachim [...]. 2024. *Sustainable Production and Consumption*.

global warming La dinámica de la expansión mundial de la quínoa. Bazile, Didier. 2015. *Tierra Adentro*.

Concept search

- Descripteurs ou Entités nommées
- Multi-langue, synonymes
- Hiérarchie (termes plus spécifiques)

AGROVOC Multilingual Thesaurus

Content language English Search

Alphabetical Hierarchy Groups

phenomena > natural phenomena > climate change

PREFERRED TERM

DEFINITION

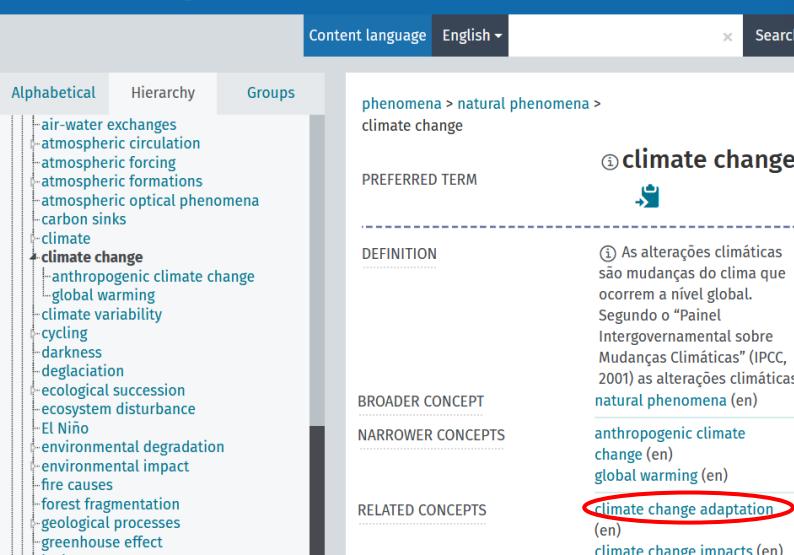
BROADER CONCEPT

NARROWER CONCEPTS

RELATED CONCEPTS

climate change adaptation (en)

climate change impacts (en)



This interface developed by the ISSA project demonstrates the enriched visualization of Agritrop documents.

Agritrop

Open Repository
of
CIRAD publications

Document
Search

Search documents

Agrovoc descriptors Wikidata named entities

Enter text and select a suggestion

climate change

Results matching only the selected descriptors

1929 result(s).

2 3 4 5 ... 242

La transition agro-écologique des agricultures du Sud. Côte, François-Xavier (ed.), Perret, Sylvain (ed.), Poirier-Magona, Emmanuelle (ed.)
[...], 2018

Climate, ca
Bengaly, Z
Perception
sahélienne
Agroforest
Bénédicte,

Results matching the selected descriptors or any more specific descriptors

4 result(s).

1

Results matching descriptors related to those selected

51 result(s).

1 2 3 4 5 6 7

climate proofing
Special issue on 'Some preliminary results of the PROCAMED project (Promotion of innovations in the camel sector for a sustainable development in The Mediterranean basin)'. Faye, Bernard. 2015. *Emirates Journal of Food and Agriculture*.



climate change adaptation
Effects of vitamin E and vitamin C on hydrogen Peroxide-induced hemolysis in moroccan dromedary camels (*Camelus Dromedarius*). Abouhafs, Rachid, Bargaa, Rita, Chakir, Youssef [...]. 2013. *Greener Journal of Medical Science*.



climate change adaptation
Projecting and valuing domestic water use at regional scale: A generic method applied to the Mediterranean at the 2060 horizon. Dumas, Patrice, Neverre, Noémie. 2015. *Water Resources and Economics*.



climate change adaptation
Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. Cissé, Ndiaga, Faye, Jacques, Foncéka, Daniel [...]. 2019. *Ecology and Evolution*.



climate change mitigation
Institutional analysis of actors involved in the governance of innovative contracts for agri-environmental and climate schemes. Barghusen, Rena, Bredemeier, Birte, Dutilly, Céline [...]. 2023. *Global Environmental Change*.



Enriched visualisation of Agritrop records

- Charte graphique d'Agritrop
- Métadonnées
- Entités nommées du résumé
- Descripteurs thématiques et géographiques
(documentalistes et calculés)
- Entités nommées géographiques



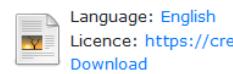
This interface developed by the ISSA project demonstrates the enriched visualization of **Agritrop** articles.

Agritrop

Open Repository
of
CIRAD publications

Oil palm cultivation in the Americas: review of the social, economic and environmental conditions of its expansion

Cifuentes-Espinosa Jaime Andrés, Feintrenie Laurene, Lesage Colombine. 2021. Oil palm cultivation in the Americas: review of the social, economic and environmental conditions of its expansion. *Cahiers Agricultures*.
<http://agritrop.cirad.fr/598930/>



Language: English

Licence: <https://creativecommons.org/licenses/by-nc/4.0/>

[Download](#)

Abstract

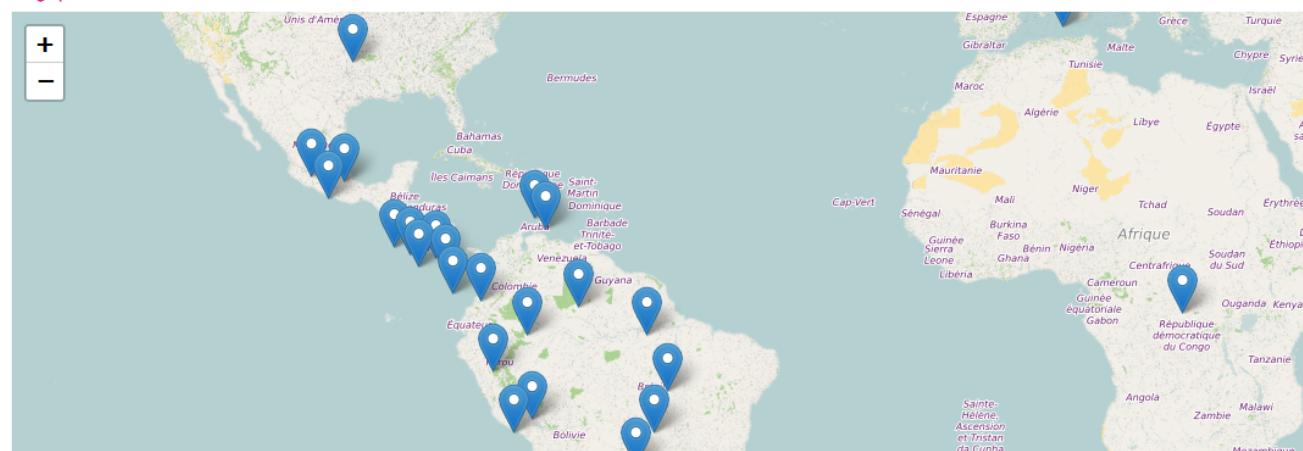
In the Americas, the **palm oil** sector has been gaining importance in the last 20 years. Although in 2018 the region only accounted for 7.1% of global **palm oil production**, it is one of the largest suitable areas for **oil palm** cultivation. We conducted a literature review on how the sector developed and how its development influenced private and public actors in their **land use** decisions. Three categories of arrangements between **oil palm** growers and **palm oil** extraction units. We grouped cases reported in the literature in three categories emerged in response to the call for better integration of sustainability in the value chain. These categories represent almost 30% of production in the region. All the **oil palm** units are pushing **contract farming** and **partnerships with local communities**. National governments intend to regulate **oil palm** production, and for **sustainable production**; they now have to face challenges. However, there are still many negative impacts on the environment, on local populations, and on **biodiversity**. Thus, **contract farming** appears to be on the way to being leaders of sustainability in the **palm oil** sector, challenges remain.

[Hide named entities](#)

Agrovoc descriptors

palm oils oil palms deforestation land use agroforestry guineensis Caribbean Central America Latin America South America Brazil Colombia sustainable agriculture environmental impact Agrovoc land use degradation socioeconomic development

Geographic named entities extracted from the text



Geographical search

le référentiel contient des termes géographiques

The screenshot displays two separate search results pages from the Agrovoc search interface.

Top Search: The search bar contains "West Africa" and "food policies". The results are titled "Results matching the selected descriptors or any more specific descriptors" and show 29 results. The first result is for Benin, followed by Côte d'Ivoire, Togo, Burkina Faso, and food policies.

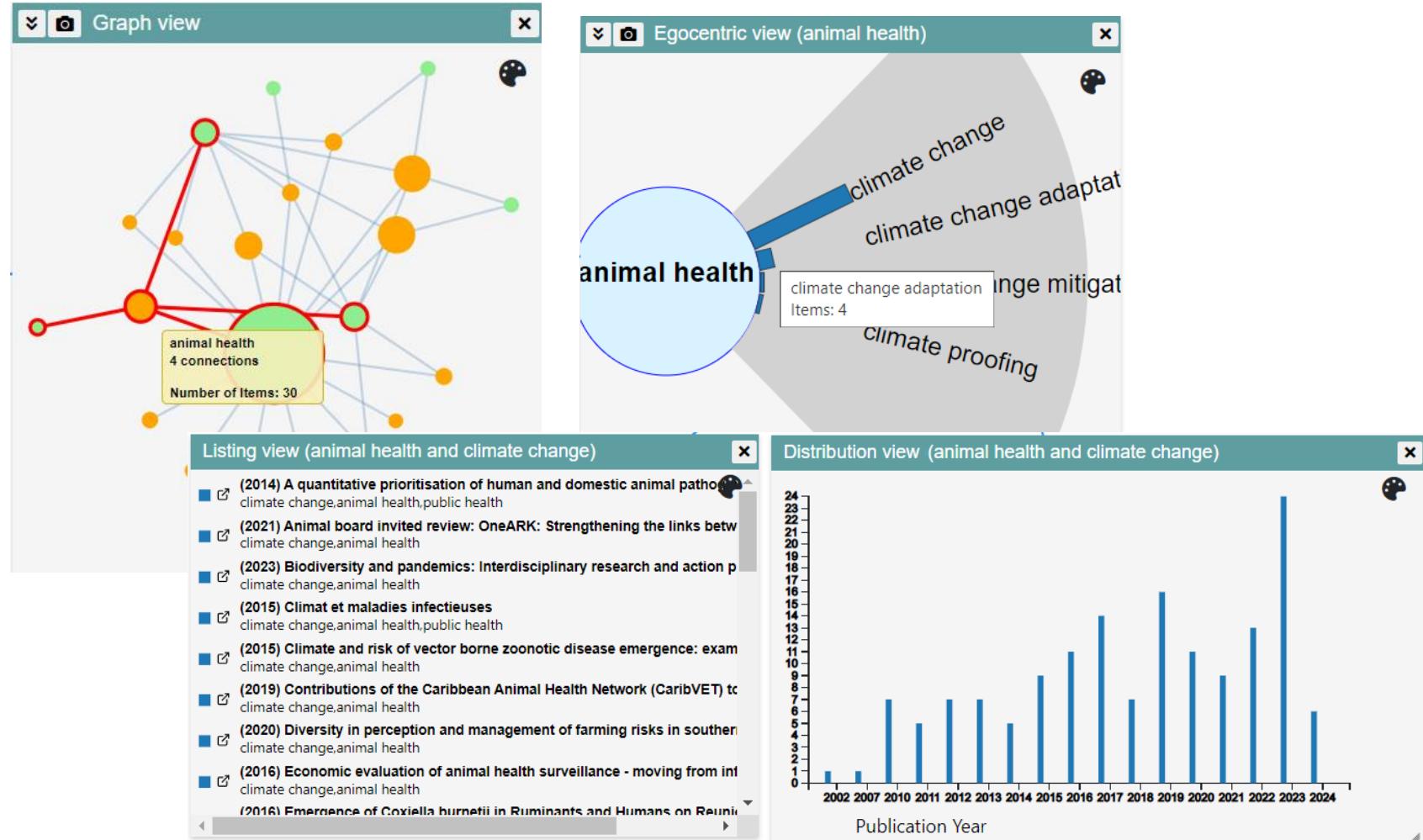
Bottom Search: The search bar contains "organic farming" and "Southeast Asia". The results are titled "Results matching the selected descriptors or any more specific descriptors" and show 6 results. The first result is for organic agriculture in Thailand, followed by Indonesia, Vietnam, Cambodia, and organic agriculture.

CollEx-Persée Logo: In the top right corner, there is a logo for CollEx-Persée featuring a magnifying glass over a map, with the text "CollEx-Persée" and "ISSA" above it.

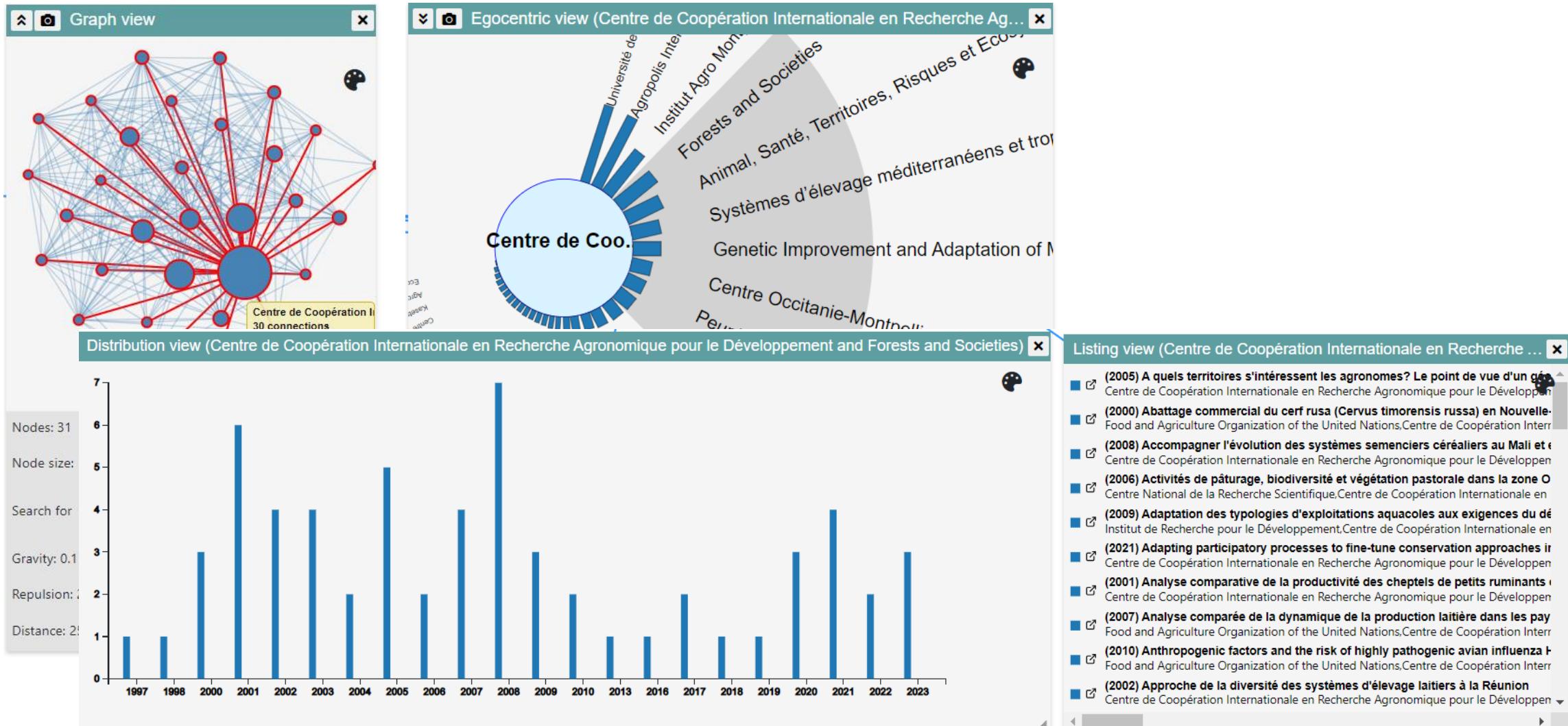
Archive exploration: articles on climate change and health ?

climate change
anthropogenic climate change
global warming

health
animal health
ecosystem health
forest health
human health
mental health
plant health
public health
rangeland health
reproductive health
seed health



Archive exploration: collaborations with other institutions



Archive exploration: temporal distribution of publications

Sustainable development goals (SDG) in OpenAlex

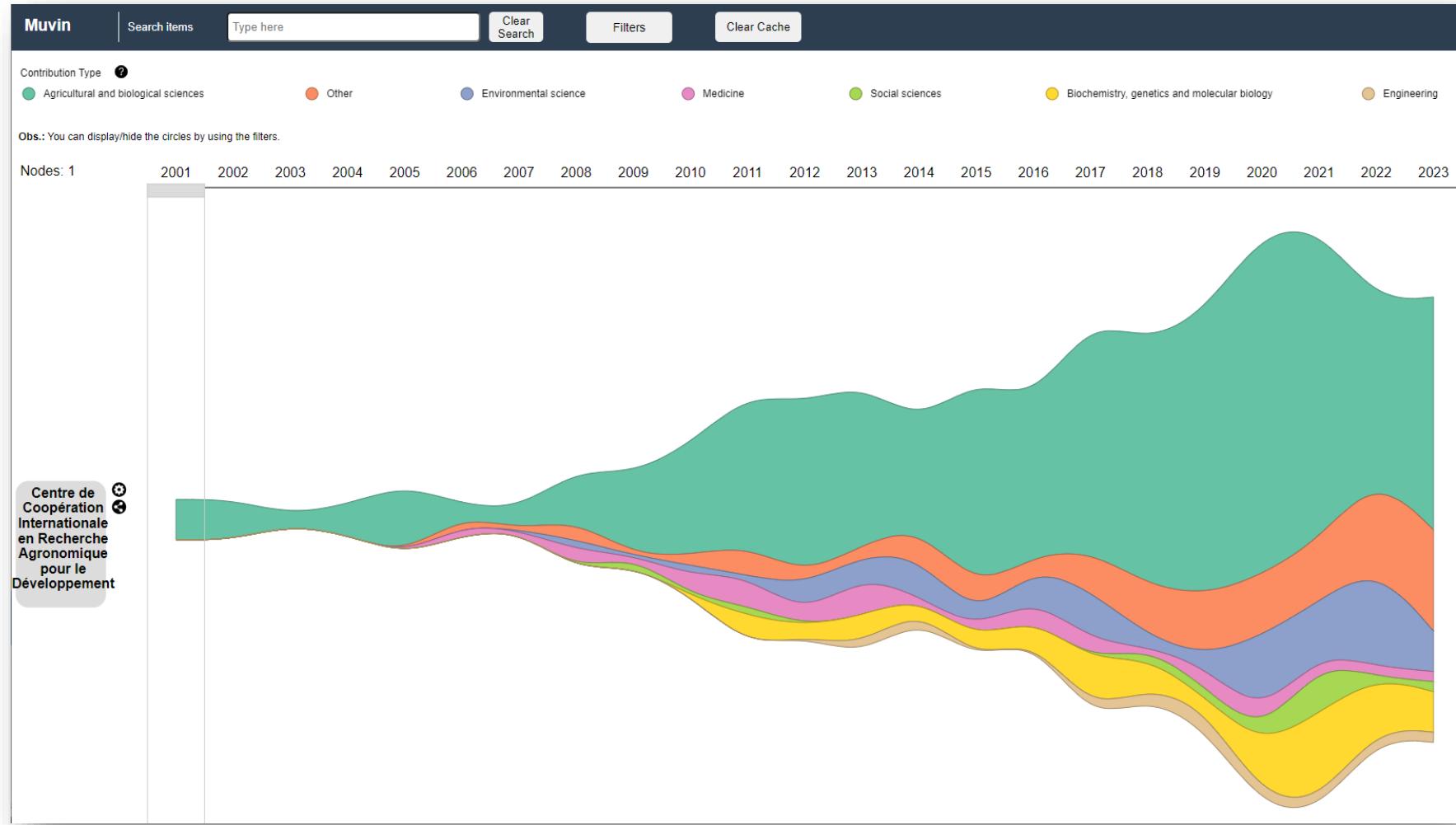
institutes



Archive exploration: temporal distribution of publications

Scientific field)
in OpenAlex

institutes



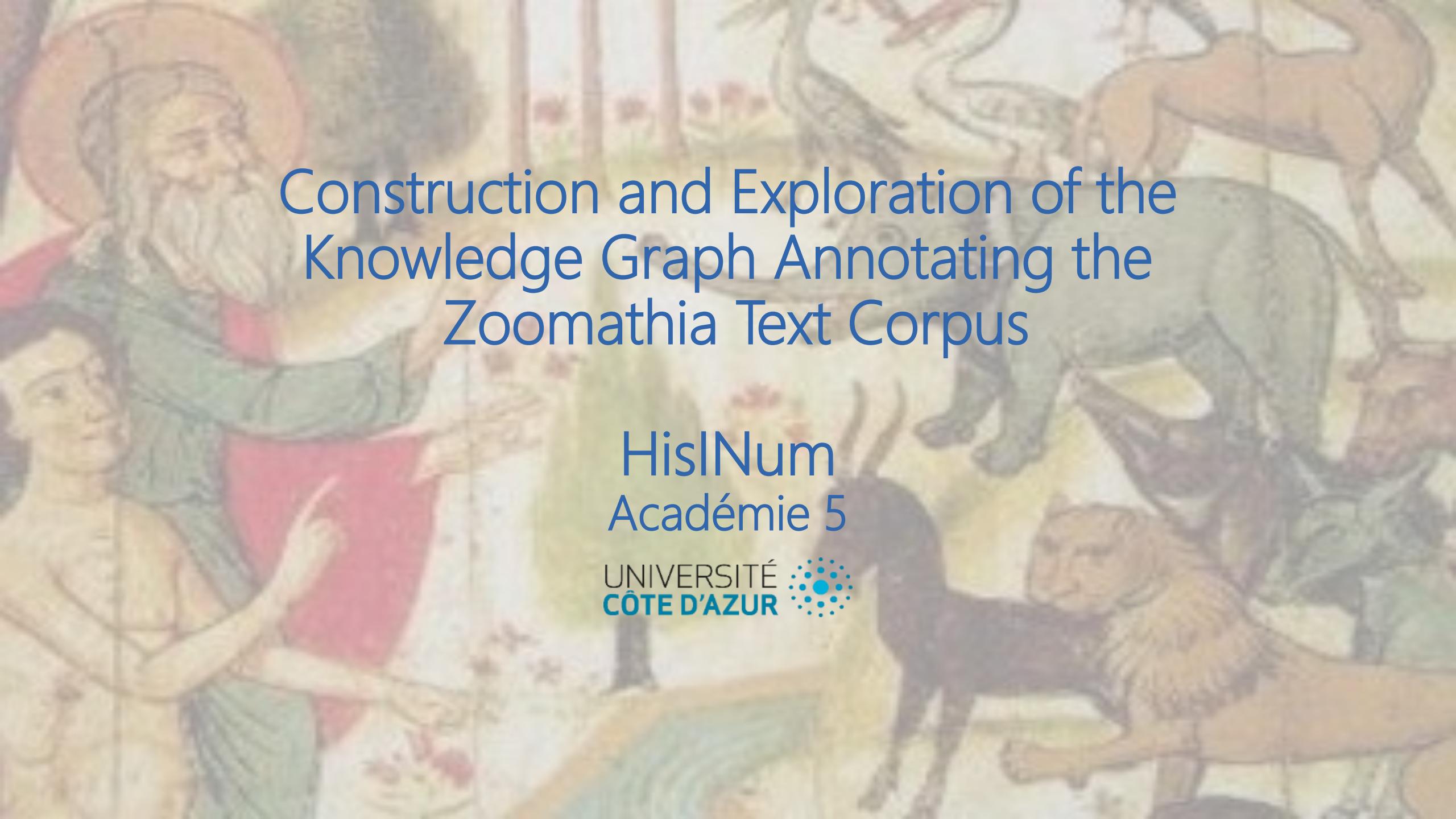
Semantic Indexing of a Scientific Archive in Agriculture



<http://issa.i3s.unice.fr/search/>

<https://dataviz.i3s.unice.fr/mgexplorer/dashboard> with SPARQL endpoint <https://data-issa.cirad.fr/sparql>

Toulet A., Michel F., Bobasheva A., Menin A., Dupré S. ISSA : un graphe de connaissances au service de la recherche bibliographique, EGC 2023



Construction and Exploration of the Knowledge Graph Annotating the Zoomathia Text Corpus

HislNum
Académie 5

UNIVERSITÉ
CÔTE D'AZUR



Semantic annotation of Historia Naturalis of Pliny the Elder

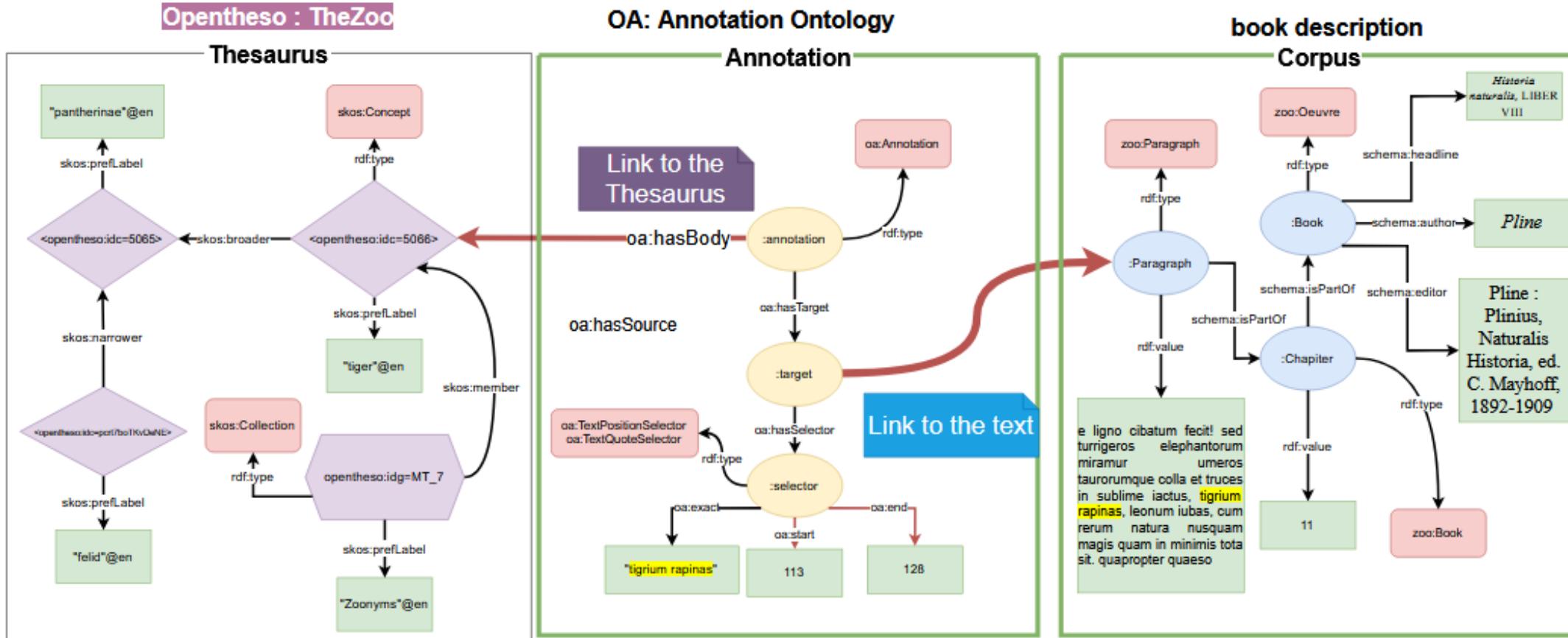
Thesaurus TheZoo

The screenshot displays the Thesaurus TheZoo application interface. On the left, a vertical tree view shows a concepts hierarchy, with 'tiger' highlighted. In the center, a detailed view of the concept 'tiger' is shown, including its classification under 'TERRESTRIAL > WILD BEAST > australasia > bilateria > deuterostoma > tetrapoda > vertebrata > mammalia > carnivora > feliformidae > feline' and its specific names in various languages. Below this is a 'Concepts Collection' section. On the right, a 'Label translations' section shows the word 'tiger' in multiple languages. Labels 'Concepts label', 'Concepts Collection', and 'Label translations' are overlaid on the interface.

Manual annotation of the text

The screenshot shows a manual annotation interface. A red box highlights a portion of a Latin paragraph from Pliny's Historia Naturalis. Annotations are overlaid on the text, such as 'sun : neck; ferocity' and 'Irene Trimegiste: tiger - ferocity'. An arrow points from a 'Paragraph' label to the text. Labels 'Paragraph' and 'Annotation' are overlaid on the interface.

Construction of the Zoomathia knowledge graph



Automatic annotation of the Zoomathia corpus

- Classifier
- Named entity recognition on DBpedia and Wikidata

DBpedia Spotlight

spaCy fishing (Entity-Fishing)

Exploration of the Zoomathia corpus

The screenshot shows a search results page for the Zoomathia corpus. At the top, there is a navigation bar with links for Home, Explore the corpus, Explore a work (which is highlighted in blue), and Competency questions. The main content area has two search fields: 'Author' (Pliny the Elder) and 'Work' (Historia naturalis). Below these are filters for 'Editor: ed. C. Mayhoff', 'Date: 1892-1909', and an 'Export' button for XML-TEI. On the left, there is a 'Table of content' for 'Book - LIBER VIII' with 17 numbered paragraphs. Two specific paragraphs are highlighted with yellow boxes: [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 ac lunae veneratio.' and [2] 'Auctores sunt in Mauretaniae saltibus ad quandam amnem, cui nomen est Amilo, nitescente luna nova greges eorum descendere ibique se purificantes sollemniter aqua circumspergi atque ita salutato sidere in silvas reverti vitulorum fatigatos prae se ferentes.' To the right, there is a section titled 'Book LIBER VIII' with a list of 15 concepts: WILD, astral cult, comparison with human behaviour and qualities, comparison with human behaviour and qualities difference in favour of animal, comparison with human behaviour and qualities similarity, elephant, language comprehension, memory, moon cult, pleasure, relative size, religion, sense of justice, size, sun cult. Below this is another section with 16 concepts: WILD, astral cult, elephant, gregarious, habitat, habitat forest, intellectual authority, moon phase, new moon, purification ritual, religion, ritual, solidarity, terrestrial habitat.

Author

Pliny the Elder

Work

Historia naturalis

Editor: ed. C. Mayhoff

Date: 1892-1909

Export: XML-TEI

Table of content

Book - LIBER VIII

[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 ac lunae veneratio.

[2] Auctores sunt in Mauretaniae saltibus ad quandam amnem, cui nomen est Amilo, nitescente luna nova greges eorum descendere ibique se purificantes sollemniter aqua circumspergi atque ita salutato sidere in silvas reverti vitulorum fatigatos prae se ferentes.

Book LIBER VIII

15 concepts

WILD, astral cult, comparison with human behaviour and qualities, comparison with human behaviour and qualities difference in favour of animal, comparison with human behaviour and qualities similarity, elephant, language comprehension, memory, moon cult, pleasure, relative size, religion, sense of justice, size, sun cult

16 concepts

WILD, astral cult, elephant, gregarious, habitat, habitat forest, intellectual authority, moon phase, new moon, purification ritual, religion, ritual, solidarity, terrestrial habitat

Exploration of the Zoomathia corpus

The screenshot shows the Zoomathia corpus exploration interface. At the top, there is a navigation bar with links for Home, Explore the corpus (which is highlighted in blue), Explore a work, and Competency questions. The main content area has a title "Define a custom filter" and a note: "This form can take multiple values for each search field. The default the result is union between values of the author and the work search field." Below this are three filter boxes: "Filter on author(s)" containing "Aelian" and "Pliny the Elder"; "Filter on work(s)" containing "Bruta animalia ratione uti"; and "Filter on concept(s)" containing "BIRD". There is also a "Search" button and a "OR AND" toggle. The results section shows "Results Number of Work: 1". On the left, there is a "Table of content" for "Oeuvre - Historia naturalis" with sections for Book - LIBER VIII and Book - LIBER IX. The right side displays the text of "Historia naturalis" from Book LIBER VIII, specifically Paragraph 36, which discusses Megasthenes' observations of Indian serpents and Metrodorus' account of deer hunting in Pontus. A sidebar on the right lists "17 concepts" related to the text, including BIRD, India, Megasthenes, Metrodorus, Place, carnivorous, deer, flying, historian, historical character, intellectual authority, locomotion, nourishment, philosopher, predation, size, and snake. Another section below lists "13 concepts" including alopecia, bear, cold, fat, hibernation, house building, human disease, incubation, medical use of animal product, and nourishment.

Zoomathia
ancient mosaics

Home Explore the corpus Explore a work Competency questions

Define a custom filter

Note: This form can take multiple values for each search field. The default the result is union between values of the author and the work search field.

Filter on author(s): Aelian x Pliny the Elder x

Filter on work(s): Bruta animalia ratione uti x

Filter on concept(s): BIRD x

OR AND

Search

Results
Number of Work: 1

Table of content

- Oeuvre - Historia naturalis
 - Book - LIBER VIII
 - Paragraph - 36 [36] Megasthenes scribit in India serpentes in tantam magnitudinem adolescere, ut solidos hauriant cervos tauros que, Metrodorus circa Rhydacum amnem in Ponto, supervolantes ut quamvis alte perniciter que alites haustu raptas absorbeant.
 - Paragraph - 127
 - Paragraph - 142
 - Paragraph - 223
 - Book - LIBER IX
 - Paragraph - 37
 - Paragraph - 63
 - Paragraph - 186 [127] Specus si non habuere, ramorum fruticum que congerie aedificant, impenetrabilis imbris molli que fronde constratos. Primis diebus bis septenis tam gravi somno premuntur, ut ne vulneribus quidem excitari queant. tunc mirum in modum veterno pinguescant. Illi sunt adipes medicaminibus apti contra que defluvium capilli tenaces. Ab his diebus residunt ac priorum pedum succu vivunt. Fetus rigentes adprimendo pectori fovent non alio incubitu quam ad ova volvres.
 - Book - LIBER X
 - Book - LIBER XI

Historia naturalis

Book LIBER VIII

Megasthenes scribit in India serpentes in tantam magnitudinem adolescere, ut solidos hauriant cervos tauros que, Metrodorus circa Rhydacum amnem in Ponto, supervolantes ut quamvis alte perniciter que alites haustu raptas absorbeant.

Specus si non habuere, ramorum fruticum que congerie aedificant, impenetrabilis imbris molli que fronde constratos. Primis diebus bis septenis tam gravi somno premuntur, ut ne vulneribus quidem excitari queant. tunc mirum in modum veterno pinguescant. Illi sunt adipes medicaminibus apti contra que defluvium capilli tenaces. Ab his diebus residunt ac priorum pedum succu vivunt. Fetus rigentes adprimendo pectori fovent non alio incubitu quam ad ova volvres.

17 concepts

BIRD India Megasthenes Metrodorus Place
carnivorous deer flying historian
historical character intellectual authority
locomotion nourishment philosopher
predation size snake

13 concepts

alopecia bear cold fat hibernation
house building human disease incubation
medical use of animal product nourishment

Exploration of the Zoomathia corpus

The screenshot shows the Zoomathia corpus interface with a search results table for a competency question.

Header:

- ZOOMATHIA logo
- Home
- Explore the corpus
- Explore a work
- Competency questions (selected)

Section Title: Select a competency question

Search Query: Which animals build a habitat?

Table Headers:

paragraph	name_animal	name_construction
-----------	-------------	-------------------

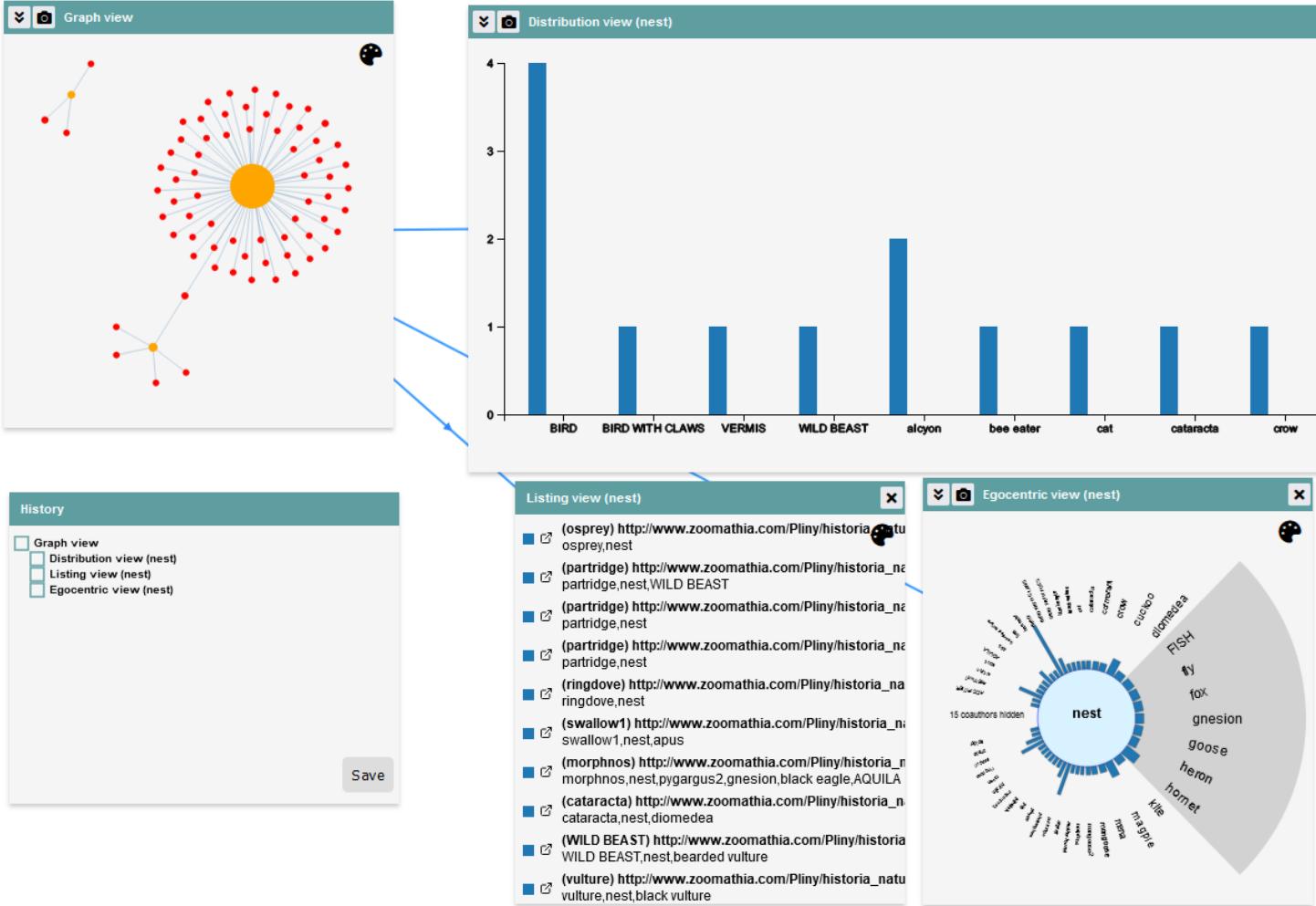
Table Data:

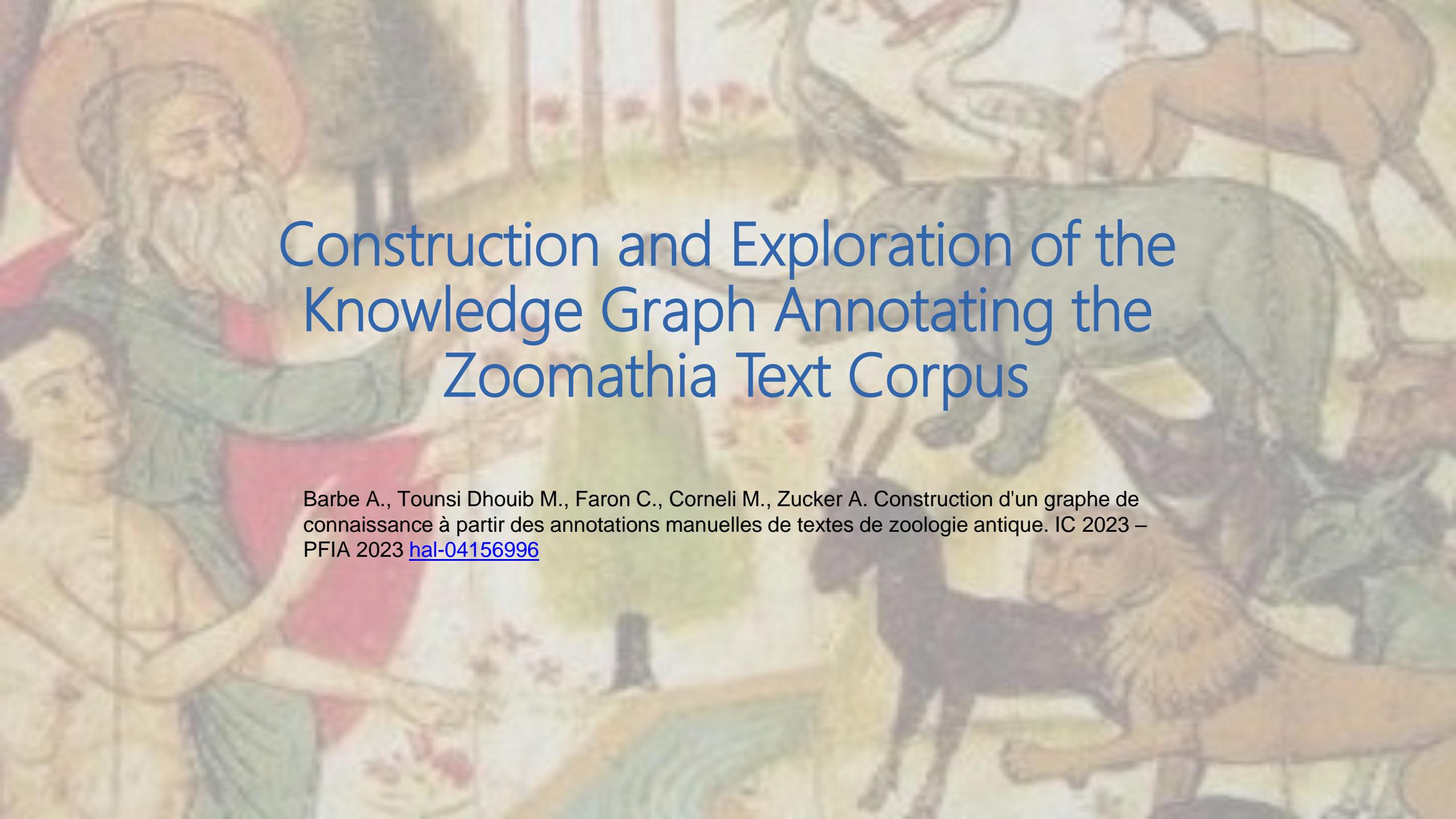
Pliny/historia_naturalis/8/218	ferret	burrow
Pliny/historia_naturalis/8/218	rabbit	burrow
Pliny/historia_naturalis/8/218	hare	burrow
Pliny/historia_naturalis/11/14	bee	hive
Pliny/historia_naturalis/11/15	bee	hive
Pliny/historia_naturalis/11/16	bee	hive
Pliny/historia_naturalis/11/22	bee	hive
Pliny/historia_naturalis/11/22	worker bee	hive
Pliny/historia_naturalis/11/23	bee	hive
Pliny/historia_naturalis/11/24	bee	hive

Pagination: Showing 1 to 10 of 92 results

Page Navigation: Previous | 1 | 2 | 3 | ... | 10 | Next

Exploration of the Zoomathia corpus





Construction and Exploration of the Knowledge Graph Annotating the Zoomathia Text Corpus

Barbe A., Tounsi Dhouib M., Faron C., Corneli M., Zucker A. Construction d'un graphe de connaissance à partir des annotations manuelles de textes de zoologie antique. IC 2023 – PFIA 2023 [hal-04156996](https://hal.archives-ouvertes.fr/hal-04156996)

Conclusion: a unified (reusable) Knowledge Graph based approach

- Semantic annotation of a text corpus
- Construction of a Knowledge Graph
- Competency questions implemented by queries on the KG
 - Concept search
 - Visualisation of enriched texts
 - Graph exploration (MGExplorer)
 - Temporal analysis (Muvin)

A Knowledge Graph based approach not limited to text annotation

- WeKG-MF: a Knowledge Graph of Observational Weather Data
- CoffeeWKG: a Weather Knowledge Graph for Coffee Regions in Colombia
- WheatObservationsKG: a Knowledge Graph of Crop Monitoring Data
- WoodKG: a Knowledge Graph of Wood and Charcoal Observations
- ...

→ Federation of complementary Knowledge Graphs:
Scientific Literature KG + Observation data KG
e.g. scientific literature on Wheat + crop monitoring data + weather data