

Ontology Engineering and Formalization: Interoperability Applied to the Biodiversity of the Kisantu Botanical Garden

Mpemba Ngoma Luz¹, Mbikayi Mpanya Jean Marcel², Mudibu Wa
Kabangu Justin³, Ikolongo Befembo Jean-Pierre⁴

¹Department of Computer Science and Technology, Higher Pedagogical Institute of Mbanza
Ngungu/ISP/DRC

²Department of Computer Science and Management, Higher Institute of Statistics of Kinshasa/ISS/DRC

³Department of Agronomy and Environment, National Pedagogical University/UPN/DRC

⁴Department of Mathematics, Statistics and Computer Science, National Pedagogical
University/UPN/DRC

Abstract

Biodiversity conservation and management in Central Africa require the formal, rigorous, and interoperable structuring of botanical knowledge. This article presents the design, mathematical formalization, and implementation of the MKBio ontology, developed for the Kisantu Botanical Garden using the METHONTOLOGY methodology [2]. Drawing on First-Order Predicate Logic, the Skolemization mechanism, and Description Logic () under the language [1,7], we formalize taxonomic, morphological, and ecosystem entities as a semantic knowledge graph. Aligned with the international Darwin Core standard [6], this model provides a deterministic symbolic foundation that serves as a safeguard against hallucinations in Neuro-Symbolic Artificial Intelligence architectures () [4,5].

Keywords: Ontology Engineering, Semantic Web, Modeling, Kisantu Botanical Garden, Description Logic, Skolemization, Darwin Core, Neuro-Symbolic AI, Graph-RAG.

1. Introduction

In the Democratic Republic of the Congo, the fragmentation of taxonomic data, the heterogeneity of field observations, and the lack of shared semantic standards hinder the scientific use and conservation of flora. Knowledge engineering [9] formally addresses this problem by translating textual descriptions into fact bases that can be computed and interpreted by any machine.

This study aims to design the MKBio ontology, a sovereign semantic foundation representing the biodiversity of the Kisantu Botanical Garden in accordance with Gruber's design principles [9] and the methodological guide by Noy and McGuinness [8].

This work focuses exclusively on ontology engineering, the rigor of logical formalizations, alignment with international standards, and integration of the model into hybrid AI architectures.

2. Formal Framework and Ontology Structuring

2. Definition and Structure

2.1. Ontology

In knowledge engineering, an ontology is a formal and explicit specification of a shared conceptualization (Gruber, 1993).

Formally, an ontology is defined as a structural 5-tuple:

$$O = \langle \mathcal{C}, \mathcal{R}_O, \mathcal{R}_D, \mathcal{I}, \mathcal{A} \rangle$$

where:

$\mathcal{C} = \{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_n\}$: the set of concepts (or classes), representing categories of entities in the domain (e.g., Plant, Animal, Medicinal Product).

$\mathcal{R}_O = \{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_m\}$: the set of relations (or properties), defined as n-ary relations over the domain. They describe class attributes and relationships between classes (e.g., hasFamily, isPollinatedBy).

$\mathcal{R}_D = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_p\}$ is the set of data properties mapping a concept to a domain of literals (, e.g. $\mathcal{R}_D \subseteq \mathcal{C} \times \mathcal{D}_L$, ex. : ScientificName).

$\mathcal{I} = \{\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_k\}$: the set of individuals (or real-world instances) belonging to the domain classes (e.g., JBK_001, Millettia laurentii).

$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$: the set of formal axioms expressing logical constraints, subsumptions, or inference rules. The axioms impose structuring constraints on $\mathcal{C}, \mathcal{R}_O, \mathcal{R}_D$ et $\mathcal{I}$. (e.g., $\text{Tree} \sqsubseteq \text{Plant}$).

2.2. Representation in Description Logic

Within the theoretical framework of the Semantic Web (), defined as a distributed knowledge graph represented as a decidable logical system guided by model theory, the ontology is translated into a knowledge base structured as a triple [2]:

$$\mathcal{K} = \langle \mathcal{I}, \mathcal{R}, \mathcal{A}_{\text{box}} \rangle$$

The (Terminological Box): Set of subsumption () and equivalence () axioms defining the conceptual schema.

The (Role Box): Set of axioms describing role hierarchies and properties (transitivity, symmetry, role disjointness).

The (Assertional Box): Set of contingent assertions about individuals, including concept assertions and role assertions [1,7].

2.3. Formal Semantics of Model Theory

The ontology's semantics are formalized by an interpretation , where represents the non-empty domain of discourse (the set of plant entities in the garden) and is an interpretation function defined by [2]:

$$\begin{aligned} \mathbf{C}^{\mathcal{I}} &\subseteq \Delta^{\mathcal{I}} \quad \forall \mathbf{C} \in \mathcal{C} \\ \mathbf{R}^{\mathcal{I}} &\subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \quad \forall \mathbf{R} \in \mathcal{R}_O \\ \mathbf{a}^{\mathcal{I}} &\in \Delta^{\mathcal{I}} \quad \forall \mathbf{a} \in \mathcal{I} \end{aligned}$$

3. Semantic Alignment with Darwin Core and Methodology

The development of the ontology follows the METHONTOLOGY life cycle (Specification, Conceptualization, Formalization, Implementation, Evaluation).

To ensure the international interoperability of the knowledge graph, an alignment function maps the set of properties to the Darwin Core standard vocabulary () [6]:

$$\begin{aligned}\Phi: (\mathcal{R}_O \cup \mathcal{R}_D) &\rightarrow \text{DwC} \\ \Phi(\text{nomScientifique}) &= \text{dwc:scientificName} \\ \Phi(\text{lieuRecolte}) &= \text{dwc:locality} \\ \Phi(\text{codeEchantillon}) &= \text{dwc:catalogNumber}\end{aligned}$$

4. Logical Formalization and Skolemization

To ensure decidability and eliminate semantic ambiguity, the fundamental taxonomic rule linking a herbarium specimen to its species and genus follows a strict transformation chain:

a) First-Order Predicate Logic :

$$\begin{aligned}\phi &\equiv \forall x \left(\text{Specimen}(x) \right. \\ &\quad \left. \Rightarrow \exists y \exists z \left(\text{Espece}(y) \wedge \text{Genre}(z) \wedge \text{aPourTaxon}(x, y) \wedge \text{appartientAuGenre}(y, z) \right) \right)\end{aligned}$$

b) Skolemized Normal Form :

The elimination of the existential quantifiers and dependent on is performed by introducing the Skolem functions and :

$$\begin{aligned}\phi_{\text{Skolem}} &\equiv \forall x \left(\neg \text{Specimen}(x) \right. \\ &\quad \vee \left(\text{Espece}(f_{\text{esp}}(x)) \wedge \text{Genre}(f_{\text{gen}}(x)) \wedge \text{aPourTaxon}(x, f_{\text{esp}}(x)) \right. \\ &\quad \left. \left. \wedge \text{appartientAuGenre}(f_{\text{esp}}(x), f_{\text{gen}}(x)) \right) \right)\end{aligned}$$

c) Projection into Description Logic [1,10]:

$$\text{Specimen} \sqsubseteq \exists \text{aPourTaxon.Espece} \sqcap \leq 1 \text{aPourTaxon.Espece}$$

$$\text{Espece} \sqsubseteq \exists \text{appartientAuGenre.Genre} \sqcap \leq 1 \text{appartientAuGenre.Genre}$$

5. Practical Examples of Skolemization in Botany

To illustrate the transition from natural-language expressions to a computable form in Protégé [10], three key rules in the botanical domain are formalized and Skolemized:

Case 1: One-variable functional dependency

Statement: Every herbarium specimen was collected by at least one botanist.

$$\text{First-Order Predicate Logic: LPO : } \phi_1 \equiv \forall x \left(\text{Specimen}(x) \Rightarrow \exists y \left(\text{Botaniste}(y) \wedge \text{aEteRecoltePar}(x, y) \right) \right)$$

$$\text{Skolemized form : } \phi_{1, \text{Skolem}} \equiv \forall x \left(\neg \text{Specimen}(x) \vee \left(\text{Botaniste}(f_{\text{bot}}(x)) \wedge \text{aEteRecoltePar}(x, f_{\text{bot}}(x)) \right) \right)$$

Case 2: Nested functions under a universal quantifier

Statement: Every botanical genus belongs to at least one family, which is associated with an order.

First-Order Predicate Logic:

$$\begin{aligned} &\equiv \forall x \left(\text{Genre}(x) \right. \\ &\quad \Rightarrow \exists y \exists z \left(\text{Famille}(y) \wedge \text{OrdreBotanique}(z) \wedge \text{appartientAFamille}(x, y) \right. \\ &\quad \left. \left. \wedge \text{rattacheAOrdre}(y, z) \right) \right) \end{aligned}$$

Skolemized form: $\phi_{2, \text{Skolem}} \equiv \forall x \left(\neg \text{Genre}(x) \vee \left(\text{Famille}(f_{\text{fam}}(x)) \wedge \text{OrdreBotanique}(f_{\text{ord}}(x)) \wedge \text{appartientAFamille}(x, f_{\text{fam}}(x)) \wedge \text{rattacheAOrdre}(f_{\text{fam}}(x), f_{\text{ord}}(x)) \right) \right)$

Case 3: Skolemization by constant (absence of a universal quantifier)

Statement: There exists at least one threatened plant species that has a medicinal use.

First-Order Predicate Logic: $\phi_3 \equiv \exists x \exists y \left(\text{Espece}(x) \wedge \text{estMenacee}(x) \wedge \text{UsageMedicinal}(y) \wedge \text{aPourUsage}(x, y) \right)$

Skolemized form: $\phi_{3, \text{Skolem}} \equiv \text{Espece}(c_{\text{esp}}) \wedge \text{estMenacee}(c_{\text{esp}}) \wedge \text{UsageMedicinal}(c_{\text{usg}}) \wedge \text{aPourUsage}(c_{\text{esp}}, c_{\text{usg}})$

7. Neuro-Symbolic Model and Safeguard Against Hallucinations (Graph-RAG)

In life sciences, generative language models () present a critical risk of factual hallucination [4,5]. To overcome this limitation, integrating the MKBio ontology into a hybrid neuro-symbolic architecture based on the paradigm will improve reliability during retrieval:

$$A = f_{\text{LLM}}(q, \text{SPARQL}(q, \mathcal{K}_{\text{KBio}}))$$

where is the natural-language query submitted by the user, and is the instruction for extracting the subgraph of certified facts from the ontology. Since the system's consistency is formally guaranteed (), the generated response is strictly deterministic, scientifically accurate, and certified as hallucination-free.

6. Concept Description

6.1. Taxonomy: Concept Dictionary (OWL Classes)

OWL Identifier	Class	Subsumption relation (rdfs:subClassOf)	Definition and ecological scope within the Kisantu Botanical Garden
JBK_Concept		owl:Thing	Universal root concept encompassing all abstract and concrete entities in the knowledge base.
Entite_Biologique		:JBK_Concept	Biological superclass grouping taxa of living organisms (fauna, flora, fungal kingdom) recorded in situ.
Vegetal		:Entite_Biologique	Central concept representing the Garden's plant population. It is the main application domain of inference rules.
Arbre		:Vegetal	Morphological specialization denoting large woody phanerophytes with a distinct trunk.

Arbuste	:Vegetal	Morphological specialization describing medium-sized woody plants branching from the base.
Liane	:Vegetal	Biological specialization grouping woody or herbaceous climbing plants that depend on physical support.
Graminee	:Vegetal	Non-woody individuals belonging to the herbaceous layer or the Poaceae family.
Regne_Animal	:Entite_Biologique	Concept encompassing local fauna acting as vectors of ecosystem services or trophic interactions.
Insecte	:Regne_Animal	Class representing entomological fauna (bees, ants, Lepidoptera) involved in reproductive dynamics (pollinators).
Mammifere	:Regne_Animal	Class representing mammalian fauna (primates, bats, rodents) involved in dispersal cycles.
Oiseau	:Regne_Animal	Class of the Garden's avifauna participating in mutualistic relationships (ornithophily and ornithochory).
Champignon	:Entite_Biologique	Organisms of the Fungi kingdom that maintain symbiotic (mycorrhizal) or parasitic relationships with roots.
Entite_Spatiale	:JBK_Concept	Geographical abstraction enabling the spatial representation of botanical data.
Localisation_JBK	:Entite_Spatiale	Actual, delimited physical sectors of the Garden (Arboretum, Palmetum, Bambusetum, Greenhouse Sector, banks of the Inkisi).
Famille	:JBK_Concept	Class grouping botanical families according to the APG (Angiosperm Phylogeny Group) phylogenetic classification.
Saison	:JBK_Concept	Chronological and climatic divisions of the Kongo-Central region (dry season, rainy season).
Statut_UICN	:JBK_Concept	Official vulnerability indicators of the International Union for Conservation of Nature (EN [Endangered], VU [Vulnerable], LC [Least Concern], NT [Near Threatened]).
Usage	:JBK_Concept	Categorization of human uses of plants. It is subdivided into UsageMedicinal and UsageAutre.

6.2. Object Properties

Object properties express relationships between individuals in the ontology. In this modeling, several properties play a central role:

Object property name	Role
aPourFamille	Links a plant to its botanical family.
aPourImpact	Links a plant to an ecological impact.
aPourLocalisation	Links a plant to its location in the Garden.
aPourOrigine	Links a plant to its geographical origin.
aPourPeriodeFloraison	Links a plant to its flowering period.
aPourStatutUICN	Links a plant to its conservation status.
aPourStrategieReboisement	Links a plant to a reforestation strategy.
aPourUsageAutre aPourUsageMedicinal	Link a plant to its uses.
estDissemine Par estPollinise Par,esthotede estSymbiote	Model ecological relationships between plants and other biological entities.
fleuritEnSaison	Specifies the flowering season.

7. Implementation Using the Protégé Tool

Protégé is an ontology development environment widely used in research on the Semantic Web, knowledge representation, and ontology engineering. It enables the creation, structuring, editing, and validation of ontologies according to RDF, RDFS, and OWL standards. In this work, Protégé served as the main tool for implementing the ontology dedicated to biodiversity information retrieval [10] [1] [2] [3][4] [7].

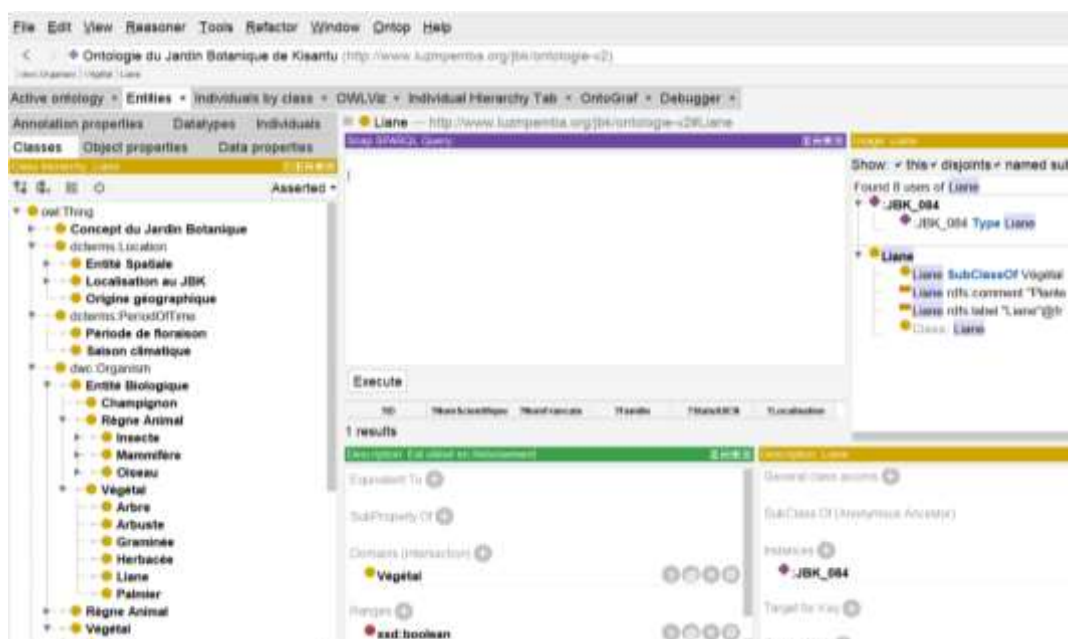


Figure 1: Protégé interface showing the implementation of our ontology with 3,797 axioms

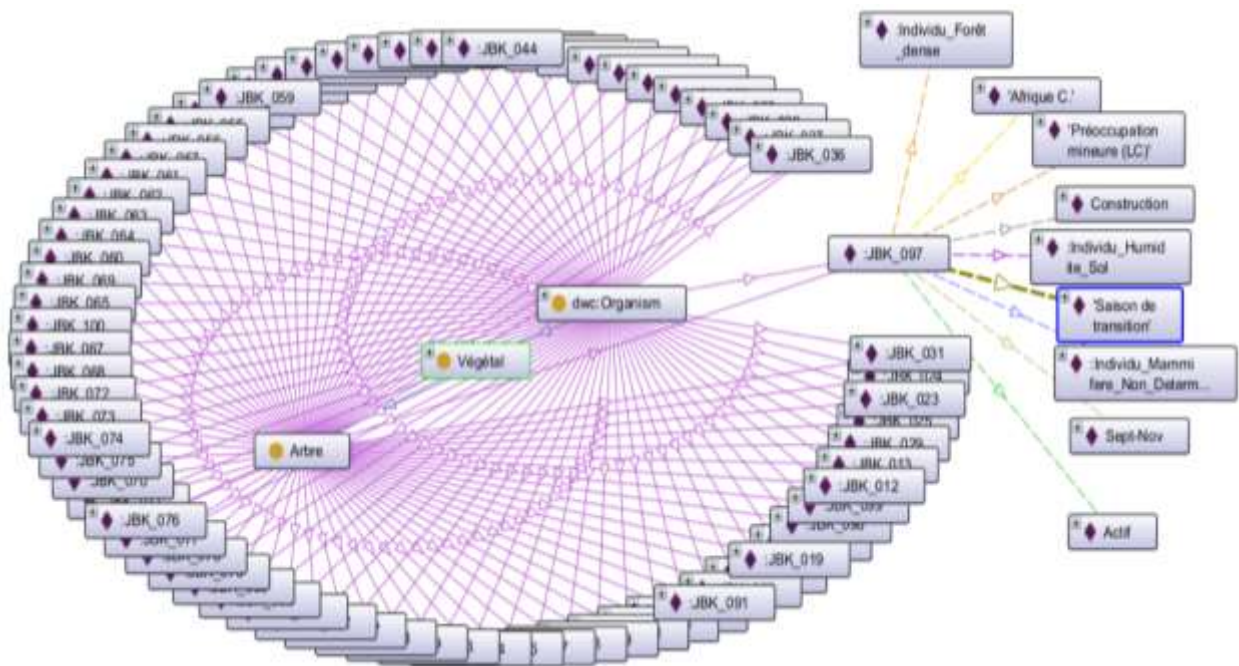


Figure 2: Illustration of our MKBio knowledge graph with 100 plant species linked to the Darwin Core vocabulary and plant entities

9. Scientific Contributions

- Heritage and taxonomic contribution: Semantic digitization and structuring of the first formal knowledge graph dedicated to species of the Kisantu Botanical Garden.
- Logical and theoretical contribution: Establishment of an uninterrupted formal chain () [1,7] ensuring the decidability of the schema.
- Contribution to international interoperability: Complete semantic alignment with the Darwin Core standard () [6], enabling the direct integration of Congolese data into international networks.
- Contribution to modern AI: Use of the ontology as a deterministic symbolic safeguard within an architecture that neutralizes hallucinations by large language models [4,5].

10. Conclusion and Perspectives

This article has demonstrated how ontology engineering and formal logic constitute major methodological levers for preserving, structuring, and enhancing Africa's botanical heritage. By establishing a rigorous and uninterrupted formalization chain, from First-Order Predicate Logic to Description Logic under the standard, through the Skolemization mechanism, this work ensures a concept schema that is theoretically sound, decidable, and free of semantic ambiguity.

MKBio goes beyond heritage preservation alone, positioning itself as a deterministic symbolic barrier against contemporary generative artificial intelligence.

This study places the Kisantu Botanical Garden at the forefront of digital biodiversity in Central Africa. It paves the way for extending this semantic infrastructure to other ecosystems and protected areas of the Congo Basin, laying the foundations for an integrated national botanical knowledge network serving scientific research, sustainable conservation, and digital sovereignty.

References

1. Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., & Patel-Schneider, P. F. (Eds.). (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
2. Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.
3. Glimm, B., Horrocks, I., Motik, B., Shearer, R., & Stoilos, G. (2014). HermiT: An OWL 2 reasoner. *Journal of Automated Reasoning*, 53(3), 245-269.
4. Garcez, A. d., & Lamb, L. C. (2023). Neurosymbolic AI: The 3rd Wave. *Artificial Intelligence Review*, 56(11), 12387-12406.
5. Pan, J. Z., Vetere, G., Gomez-Perez, J. M., & Wu, H. (Eds.). (2021). *Explaining Linked Data: Integration of Knowledge Graphs and Machine Learning*. Springer.
6. Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., ... & Vieglais, C. (2012). Darwin Core: An evolving community-developed standard for biodiversity data. *PLoS ONE*, 7(1), e29715.
7. W3C OWL Working Group. (2012). *OWL 2 Web Ontology Language Document Overview (Second Edition)*. W3C Recommendation.
8. Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05.
9. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
10. Horridge, M. (2011). *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools (Edition 1.3)*. The University of Manchester.