

Ontology-driven taxonomic workflows for Afrotropical Bees

Aurona Gerber^{a,b}, Nishal Morar^{a,c}, Thomas Meyer^{a,d}

^aCenter for Artificial Intelligence Research (CAIR), Cape Town, South Africa

^bDepartment of Informatics, University of Pretoria, South Africa

^cDepartment of Computer Science, University of Kwazulu-Natal, South Africa

^dDepartment of Computer Science, University of Cape Town, South Africa



Abstract

This poster presents the results of an investigation into the use of ontology technologies to support taxonomy functions. Taxonomy is the science of naming and grouping biological organisms into a hierarchy. A core function of biological taxonomy is the classification and revised classification of biological organisms into an agreed upon taxonomic structure based on sets of shared characteristics. Recent developments in knowledge representation within Computer Science include the establishment of computational ontologies. Such ontologies are particularly well suited to support classification functions such as those used in biological taxonomy.

Using a specific genus of Afrotropical bees, this research project captured and represented the taxonomic knowledge base into an OWL2 ontology. In addition, the project used and extended available reasoning algorithms over the ontology to draw inferences that support the necessary taxonomy functions, and developed an application, the *web ontology classifier* (WOC). The WOC uses the Afrotropical bee ontology and demonstrates the taxonomic functions namely: identification (keys) as well as the description and comparison of taxa (taxonomic revision).

The Problem: Taxonomic Impediment

The *taxonomic impediment* is the term used to describe the crises in taxonomy due to a shortage of skills and resources, as well as a lack of funding and interest by younger scientists [6]. Taxonomy is under pressure due to high demand for identification services, as well as the lack of expertise and resources due to the highly specialised and time-consuming functions taxonomists perform.

Using the specific case under investigation namely Afrotropical Bees, very few taxonomists worldwide exist, and most of them are considering retirement. Afrotropical bees are one of the largest pollinator groups in Southern Africa, which makes them a biologically important group of organisms [3, 4] and the number of taxonomists that specialise in this specific group are declining significantly.

We worked with a domain specialist, Dr Connal Eardley, in this project. This research investigated how ontology technologies can assist and support biological taxonomy and taxonomists, first in describing species and their relationships to one another, and second, in the identification of specimens.



Figure 1: *Colletes capensis* Cameron A: Female; B: Male.

Taxonomists have two main functions: the identification and description of taxa, and the identification and establishment of new taxa into biological science [6]. Taxonomic revision is the description, identification and/or revision of new genera or species based on a combination of morphological characteristics.

Taxonomists often publish taxonomic knowledge in a way that allows for identification of organisms namely a key. Taxonomic keys are tools constructed from taxonomic knowledge used for the identification of organisms or taxa using unique characteristics or features [12]. These features represent the set of characteristics that can uniquely describe a taxa or taxon. The keys display the features as choices to the user, and through the choices, the user navigates the set of features until a specific taxon is identified. Keys are often computerized by using platforms such as Lucid [8, 9].

Male. Length 9.3–11.1 mm; integument usually black, T6–T7 orange; clypeus gently convex, short, clypeo-ocellus:clypeus 1:0.7; antennal flagellum convex ventrally; vertex short, orbit-ocellus 1:1.0; orbit-ocellus:gena length 1:1.0; mandible with three similar teeth, a large tubercle medially on ventral edge; forecoxa spinose; tarsi unmodified, without large fringes; T3–T5 with distinct preapical ridges.

Identification key
1. Female clypeus distinctly modified, ventral edge concave mediolaterally, pointed or tuberculate medially; male unknown 2
2. Female clypeus unmodified, ventral edge entire and horizontal 4
3. Ventrolateral region of clypeus flat; vestiture of T1 white <i>Megachile trichoma</i> Friese.
4. Ventrolateral region of clypeus tuberculate; vestiture on T1 black or white 3
5. Ventrolateral region of clypeus strongly tuberculate; vestiture on T1 black <i>Megachile cornigera</i> Friese
6. Ventrolateral region of clypeus slightly convex; vestiture on T1 white <i>Megachile braunsiana</i> Friese
7. Metasoma clothed mostly with bright orange vestiture 5
8. Metasoma vestiture partly white 6
9. Head and mesosoma brownish-orange <i>Megachile dorsata</i> Smith
10. Head and mesosoma mostly black <i>Megachile cognata</i> Smith
11. Metasomal dorsum with incomplete white distal fringes, black vestiture mesally, white hair semi-erect <i>Megachile lanthoptera</i> Smith
12. Metasomal dorsum mostly with reddish-orange vestiture; male T6 acutely pointed posteromedially 7
13. Mesosomal dorsum with whitish vestiture, if yellowish-brown metasomal distal fringes concolorous with mesosoma; male T6 concave posteromedially <i>Megachile discolor</i> Smith
14. Metasoma entirely clothed with white vestiture; large species, 25 mm long <i>Megachile rufoscapa</i> Friese
15. Metasomal terga black anteriorly, pallid posteriorly; small to medium sized bees, less than 19 mm long 9
16. Female T6 black; occurring in the South-western Cape Province <i>Megachile serrula</i> sp. n.
17. Vestiture on female T6 at least partly white; occurring north of the Vaal River (South Africa) 10
18. Female T6 black, male T6 white <i>Megachile angulata</i> Smith
19. Female and male T6 yellowish to orange <i>Megachile seawaldi</i> Strand

Figure 2: An example of a taxonomic single access (dichotomous) key.

Application Development

The application developed to support taxonomy is called the WOC (the Web Ontology Classifier) and it consists of three components: the Afrotropical bee ontology, the reasoning services and the web user interface.

The Afrotropical bee Ontology

Computational ontologies is a knowledge representation technology particularly suitable to support classification functions such as those fundamental to biological taxonomy. Computational ontologies are developed in the Web Ontology Language (OWL2 [7]) that is based on description logics (DLs). DLs are often described as decidable fragments of first-order logic [2]. We used the ontology editor Protégé [10] that is distributed with packaged reasoners such as Fact++ [5, 11] to develop an OWL2 ontology of the taxonomic knowledge base of Afrotropical Bees. The ontology uses the core concept diagnostic feature to represent morphological characteristics [1]. A specific bee taxon would have body parts that have diagnostic features as depicted in Figure 3.

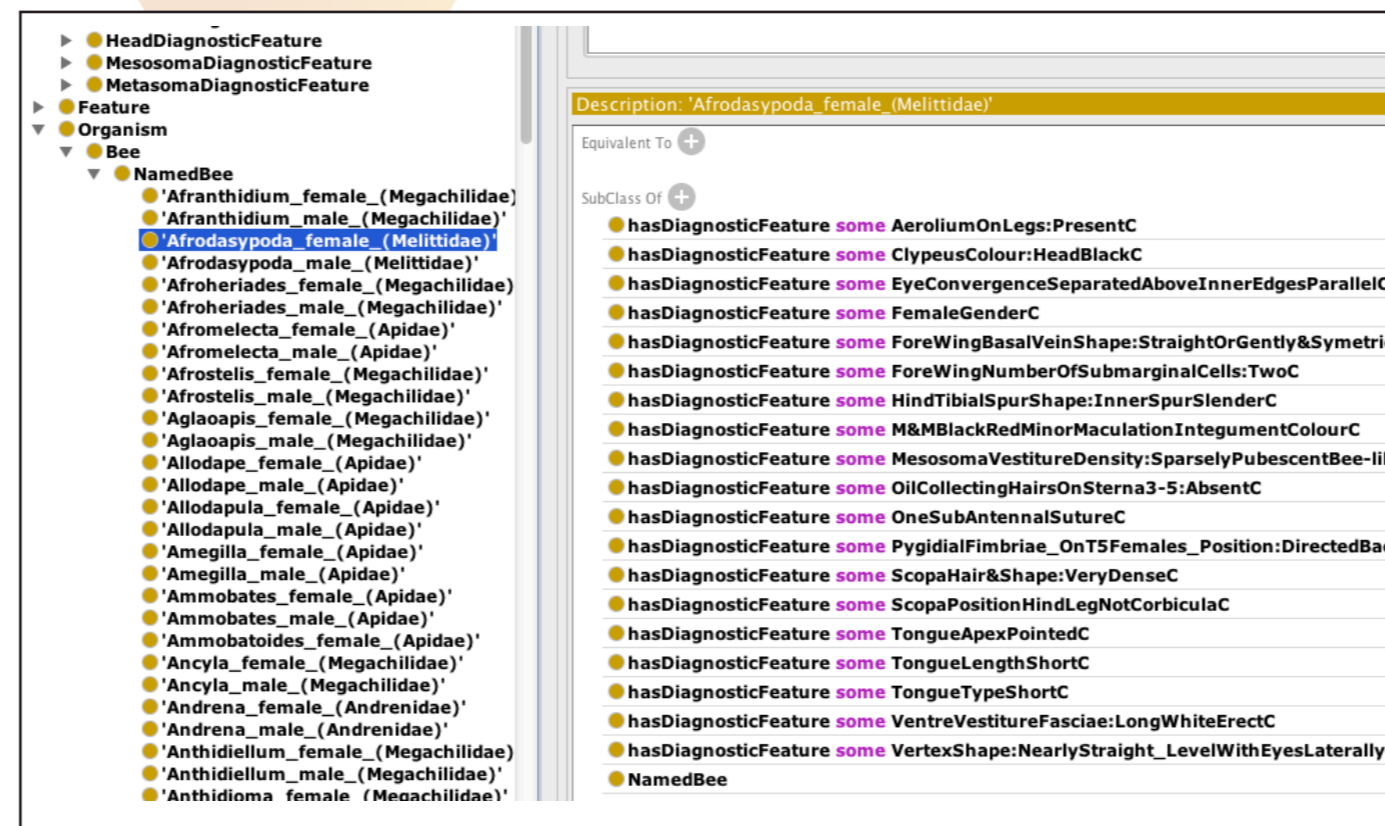


Figure 3: The genus *Afrodasyroda* with its associated diagnostic features.

The Reasoners

Because an OWL 2 ontology consists of axioms or sets of assertions based on logic, reasoners can be used to infer consequences and derive implicit knowledge from the set of explicitly stated assertions. Standard reasoners were used over the Afrotropical bee ontology for taxonomic identification services or the key functionality that can be represented by the query 'which genera or species exist that have this selected set of diagnostics features?' This query would return a set of taxa that have the selected diagnostic features.

Taxonomic revision could be represented by the query 'what set of diagnostic features does a set selected taxa exhibit and in this set, which of the diagnostic features are common to all the taxa in the set of genera?' Standard reasoning services as packaged with Protégé do not support these types of queries, also termed queries using the fillers of existential axioms and to implement this taxonomic functionality, the standard reasoning had to be complemented with reasoning supporting fillers of existential axioms.

The Web Ontology Classifier Application

The ontology and the reasoners are standardized and reusable components. Given the requirements of taxonomists, these components were integrated into a web-based application namely the *web ontology classifier* (WOC). The WOC uses the developed Afrotropical bee ontology to support the taxonomic functions *identification* and *revision* given the taxonomy of Afrotropical bees and was developed using Play Frameworks and Java.

The WOC launches with a home page with basic instructions on how to use the application. 'Classify and Query' allows for uploading an Afrotropical bee ontology. Upon submission the ontology is classified and the taxonomic information is displayed as in Figure 4.

Once classified the ontology will be displayed on the page shown in Figure 4. The diagnostic features are listed in the first column on the left, and the bee taxa modeled in the ontology are listed in the second column on the right.

The user has two options:

- select a set of diagnostic features from the first column and submit (to do an identification query), or,
- select a set of bee taxa on the right and submit (doing a revision query).

Figure 4 depicts the interface:

- Blue in the left column are the user selected (two) diagnostic features;
- Four bee taxa were found that has the selected diagnostic features (blue in the right column);
- Red in the left column are the diagnostic features that are *common* to the bee taxa;
- Green in the left are the set of *uncommon* diagnostic features associated with at least one of the identified bee taxa, but which are not common to all the bee taxa;
- Black are the *remainder* of the diagnostic features not associated with any of the selected bee taxa.

The user can now refine the selections by either adding or removing bee taxa or diagnostic features, and the two panes are updated accordingly.

The classifier is available online at <http://41.185.28.217:9000/>. The source code as well as the ontology and instructions to setup the application offline can be found on github7 - <https://github.com/Nish01/OntologyClassifierApplication>.

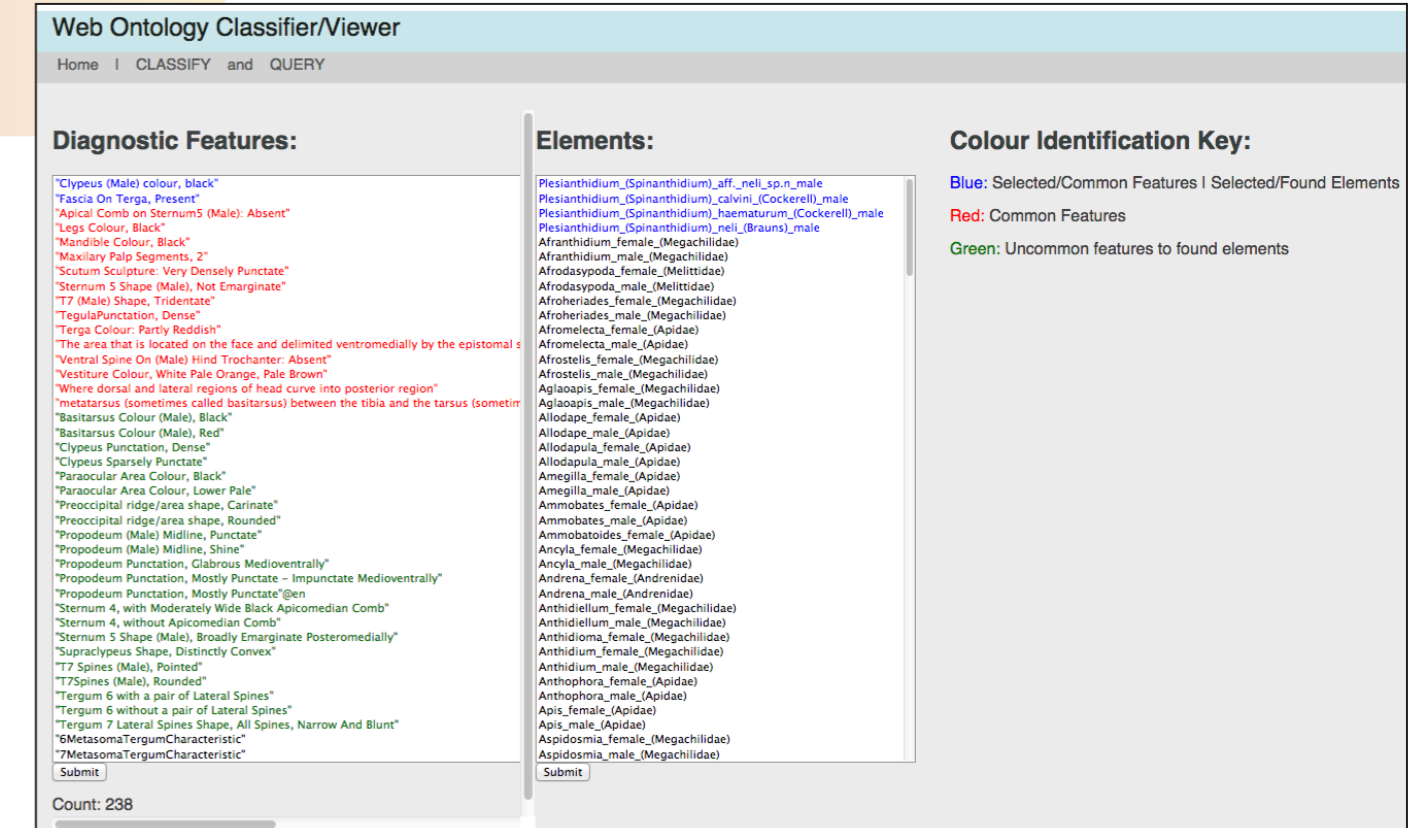


Figure 4: Two diagnostics features (blue) selected in the left hand pane result in identifying the bee taxa in blue on the right.

The additional features these blue taxa on the right all have in common are indicated in red on the left, and additional features that at least one of the taxon's are associated with are depicted in green (the uncommon features).

Conclusion

Using the specific case of Afrotropical bees, this research study had as goal investigating whether ontology technologies and ontology-based applications could support taxonomic functions. Systematics is in essence a classification system and computational ontologies are particularly well suited to resolve classification challenges, especially using qualitative data. The study developed three contributions namely: 1) the capturing and representation of the taxonomic knowledge base of Afrotropical Bees in an OWL2 ontology; 2) using and extending the use of available reasoners to draw inferences that support the necessary taxonomy workflows; and 3) the development of an application, the WOC that integrates the ontology and reasoning services. The WOC application supports taxonomic functions namely identification (keys) and taxonomic revision.

The WOC support the requirements captured from the input of the taxonomist domain expert, and also received a positive evaluation from a taxonomists in a workgroup. Further research include: 1) the development of a dedicated user interface allowing taxonomists to extend their own specific taxonomy ontologies; 2) integrating mechanisms for supporting the standardised procedures for the publishing of taxonomic revisions into the application; 3) an extended user interface for ontology editing; as well as 4) refinement of the ontological modeling, for instance including upper ontologies and existing associated domain ontologies to support standardisation and re-use of taxonomic knowledge bases.

Acknowledgement

We acknowledge the contribution of Dr. Connal Eardley, the Afrotropical bee taxonomist that provided the data and input necessary for this research study, as well as the funding provided by CAIR, the Center for Artificial Intelligence Research, CSIR, South Africa.

References

- [1] Aurona Gerber et al. An Ontology-based Taxonomic Key for Afrotropical Bees.
- [2] Baader, F. et al. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- [3] Eardley, C. et al. 2010. The Bee Genera and Subgenae of sub-Saharan Africa. *ABC Taxa*. 7 (2010).
- [4] Eardley, C. and Urban, R. 2010. Catalogue of Afrotropical bees (Hymenoptera: Apoidea: Apiformes). *Zootaxa*. 2455 (2010), 1–548.
- [5] Fact++: <http://owl.man.ac.uk/factplusplus/>.
- [6] K. Elaine Hoagland 1996. The taxonomic impediment and the convention on biodiversity. *Association of Systematics Collections Newsletter*. 24, 5 (1996), 61–62.
- [7] Krötzsch, M. 2012. *OWL 2 Profiles: An introduction to lightweight ontology languages*. Springer.
- [8] Lucid: <http://www.lucidcentral.com>.
- [9] Norton, G. et al. 2012. LucID: A multimedia educational tool for identification and diagnostics. *International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International)*, 4, 1 (2012).
- [10] The Protégé Ontology Editor: <http://protege.stanford.edu>.
- [11] Tsarkov, D. & Horrocks, I. 2006. Fact++ description logic reasoner: System description. *Automated reasoning*. Springer Berlin Heidelberg. 292–297.
- [12] Walter, D.E. & Winterton, S. 2007. Keys and the Crisis in Taxonomy: Extinction or Reinvention? *Annual Review of Entomology*. 52, 1 (Jan. 2007), 193–208.