Ontology-based Support for Taxonomic Functions

Aurona Gerber[a,b], Nishal Morar[b,d], Thomas Meyer[b,c] Connal Eardley[d,e*]

[a]*Department of Informatics, University of Pretoria, Pretoria, South Africa*
[b]*Center for Artificial Intelligence Research (CAIR), Cape Town, South Africa*
[c]*Department of Computer Science, University of Cape Town, South Africa*
[d] *Department of Computer Science, University of Kwazulu-Natal, South Africa*
[e]*Argricultural Research Council, South Africa*

Abstract

This paper reports on an investigation into the use of ontology technologies to support taxonomic functions. Within the broader context of the life sciences support for taxonomy is imperative based on several recent discussions and publications that voiced concern over the taxonomic impediment. *Taxonomy* is defined as the scientific classification, description and grouping of biological organisms into hierarchies based on sets of shared characteristics, and documenting the principles that enforce such classification. Under *taxonomic functions* we identified two broad categories: the classification functions concerned with identification and naming of organisms, and secondly classification functions concerned with categorization and revision (i.e. grouping and describing, or revisiting existing groups and descriptions).

Ontology technologies within the broad field of artificial intelligence include computational ontologies that are knowledge representation mechanisms using standardized representations that are based on description logics (DLs). This logic base of computational ontologies provides for the computerized capturing and manipulation of knowledge. Furthermore, the set-theoretical basis of computational ontologies ensures particular suitability towards classification, which is considered as a core function of systematics or taxonomy.

Using the specific case of Afrotropical bees, this experimental research study represents the taxonomic knowledge base as an ontology, explore the use of available reasoning algorithms to draw the necessary inferences that support taxonomic functions (identification and revision) over the ontology and implement a Web-based application (the WOC). The contributions include the ontology, a reusable and standardized computable knowledge base of the taxonomy of Afrotropical bees, as well as the WOC and the evaluation thereof by experts.

Keywords: taxonomy; systematics; taxonomic functions; systematics; computational ontologies; reasoning; classification; Afrotropical Bees.

---

[*] Corresponding author: *E-mail address:* aurona.gerber@up.ac.za .

Ontology-based Support for Taxonomic Functions

1. Introduction

Biodiversity is considered to be under serious threat as a result of human activities, population growth and resource consumption, including habitat conversion and urbanization, invasive alien species and over-exploitation of natural resources [1]–[3]. A disturbing fact documented is the loss of species not yet described. It is estimated that fewer than two million of an estimated 10-15 million species have been scientifically described and approximately 86% of the existing species on earth and 91% of species in the ocean still await description [4], [5]. At the core of this description of species is the practice of systematics or taxonomic classification that has been in existence for around 250 years describing approximately 1.2 million species [4].

Systematics and taxonomic classification is defined as the scientific classification, description and grouping of biological organisms into hierarchies based on sets of shared characteristics, including the identification and documentation of the principles that enforce such classification [6]. Organisms are grouped together into groups or taxa based on the Linnaean Taxonomy according to a number of shared and distinct features, usually morphological characteristics [7].

Since taxonomists are able to identify species, they play a significant role in biodiversity management including the identification of threatened species or habitats, as well as exotic pests and disease organisms. Taxonomists also offer expertise to other sectors within biology such as determining behavioral properties and patterns of species and their interactions with ecology, which could prove crucial for the continued sustainable use of natural resources [8]. The *taxonomic impediment* discussed in several publications describe the shortage of skills and resources, as well as a lack of funding and interest by younger scientists [8]. This crisis has elevated discussions about computerized support for taxonomy and the lack thereof [3], [6], [9], [10]. Most major museums house natural collections with actual data and observation records spanning decades, which provide invaluable information for biodiversity conservation [11]. These collections often also contain unstudied specimens and new species not yet described [8]. Because of the legacy of natural collections, taxonomic classifications still mostly consist of manual tasks and taxonomic data is often still paper based or in the process of being digitized [12], [13]. The wealth of information contained in the digital collections such as those of natural history museums support the urgency of extending computerized support for taxonomic functions, which provides the context for the broader research project reported on in this paper.

Taxonomists perform functions such as the identification and description of taxa, as well as the identification and establishment of new taxa into biological science. Typically, once an organism or a certain taxonomic group is identified, the next undertaking is to establish how this group can be distinguished from other taxa or other groups, and what its unique characteristics are. Under *taxonomic functions* we identified two broad categories: firstly the classification functions concerned with identification and naming of organisms, and secondly classification functions concerned with categorization (i.e. grouping and describing, or revision that is concerned with revisiting existing groups and descriptions). The thorough process that forms part of the practice termed *taxonomic revision* involves the description, identification and/or revision of groups or taxa [14]. Taxonomic revision procedures specifically are quite substantial and intense, requiring long uninterrupted hours of work by taxonomists because of the careful comparison and analysis necessary to identify and describe specific taxonomic groups.

Modern technological developments influence most scientific disciplines, and the taxonomic impediment specifically calls for computerized support for taxonomic functions. The goal of the study reported on in this paper is to determine whether the use of computational ontologies and ontology technologies could support the two major categories of taxonomic functions identified. Computational ontologies are particularly well suited for classification and categorization of concepts based on qualitative descriptions, which is a core feature of taxonomic functions. We conducted an experimental research study using the specific case of taxonomic revisions required for Afrotropical bees. We captured the existing taxonomic knowledge in an OWL ontology, used and extended the basic reasoning functionality that exists for such ontologies, and integrated these components into a Web-based application to support identification and taxonomic revision. The results obtained promise several advantages, including a standardized and reusable knowledge base for taxonomic knowledge of Afrotropical bees.

The remainder of this paper is structured as follows. Section 2 presents background on taxonomy, ontologies and ontology technologies, the available reasoning applications that could be used to draw inferences from a knowledge base, as well as ontologies within the bio-sciences. Section 3 presents the case used for this study namely the taxonomy of Afrotropical bees. In Section 4 the experimental implementation is discussed, including the ontology development, the extension of the reasoning and the development of the Web Ontology Classifier application (the WOC). Section 5 discusses the evaluation, results, contributions and implications. Section 6 concludes this study and explores possibilities and extension of this work into similar domains.

2. Background

Discussions about the computerized support for taxonomy became more prevalent since the coining of the term *taxonomic impediment* at the Convention on Biodiversity in 1995 referring to the gaps in taxonomic knowledge and the shortage of the appropriate skills [8]. Event though this convention was more than two decades ado, the taxonomic impediment is just as relevant today, if not more, given the rising concerns for biodiversity protection and sustainable development [15], [16]. Suggestions to alleviate the taxonomic impediment include computerized support, but more than a decade after the identification of the problem, computerized support is still limited [6], [10], [17]. Recent developments such as in artificial intelligence initiated renewed research in systematics, including automated taxon identification and similar applications for identification and pattern recognition such as presented in the collection of publications in the Special Volume of the Systematics Association [18].

Standardization of taxonomy and taxonomy languages is one aspect receiving attention from several developments and projects documented in literature. Open Nomenclature (ON) aims to establish a partly formal vocabulary of terms and signs or qualifiers in which a taxonomist may express remarks about their own material in the form of some abbreviated taxonomic expressions in biological classification and recent work aims to support ON intitiave with semantic standardisation [19], [20].

The Taxonomic Databases Working Group (TDWG), a Biodiversity Information Standards Group, is a non-profit association formed to establish international collaboration among biological database projects, specifically by focusing on the development of standards for the exchange of biological/biodiversity data [21]. Recent discussions in the literature include efforts for some form of community-wide, consensus-based,

human- and machine-interpretable language for describing phenotypes and their genomic and environmental contexts in order to assist with integration across key fields in biology, including systematics and ecology [22], [23]. Ontologies, specifically, are often used for such initiatives since an ontology is considered as *a formal, explicit specification of a shared conceptualisation* [24]. The specific use of ontologies within bio-sciences is discussed in Section 2.4.

2.1. Taxonomic Keys

Taxonomists often publish taxonomic knowledge in a more accessible format for the general public or non-expert user, specifically to assist non-expert users with the identification of taxa or specific species [9]. A common format used to assist with identification is the *taxonomic key*, which is taxonomic knowledge published as a series of statements or questions guiding a user towards the correct identification of an organism or taxa using the unique (often morphological) characteristics or features described as part of the taxonomic knowledge [25]. Taxonomic keys display such features as choices to the user, and through the choices, the user navigates the set of features until a specific taxon is identified. Two types of taxonomic keys exist, namely single access keys and multi-access keys:

- A single access key is an ordered taxonomic key limiting the user choices from the beginning to only the two or more options that describe a specific key feature of the taxon. The user would choose the appropriate feature based on scrutiny of the specimen, and is then directed to the next set of follow-up features until the organism is eventually identified. Single access keys are therefore divided into a further two types: *dichotomous*, if it presents the information as a series of paired mutually exclusive statements or two contrasting choices, and *polytomous* if it has more than 2 choices [6], [9]. Single access keys have the disadvantage that the user may get stuck during one of the steps if one of the features of the specimen is not identifiable (due to the specimen being damaged for example). Figure 1 shows an example of a single access key for a Southern African bee species in text format as extracted from *The Bee Genera and Subgenera of Sub-Saharan Africa* [26].
- Multi-access keys, on the other hand, are identification keys where the user can choose where to enter the identification tree with whichever feature(s) are available. The user can select one, or multiple key features from a list, resulting in the return of the set of specimens that have the selected features. By refining the feature selection, specimens that do not have the features are eliminated until only the specimen on hand is identified.

Computer supported keys are gaining some prominence due to their ease of use and accessibility. One of the first initiatives to standardize taxonomic descriptions for computer processing was DELTA [27], [28]. DELTA (Description Language for Taxonomy) is a standard data format that was originally developed in 1970 by Dallwitz of the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) with the intent to generate identification keys for insects. DELTA was refined through the years and was adopted as a standard for data exchange by Biodiversity Information Standards (TDWG). The Xper2 (and subsequent Web-based Xper3) platforms and tools are also examples of an initiative to support taxonomic descriptions and computer aided identification [29]. In Xper taxonomists use a standard format to enter taxonomic statements, which are then represented in a computer-supported multi-access key.

One of the most prominent examples of a software suite that support taxonomic applications and keys is Lucid. Lucid is a commercial software suite of tools that is marketed as powerful and highly flexible knowledge management software applications designed to help users with identification or diagnostic tasks

[30], [31]. Lucid offers both dichotomous and multi-access keys, and can include multi-media files that make the diagnostic keys user-friendly and accessible by non-expert users. Lucid also allows for Web deployment. In a Lucid key the user can select certain features in one panel, and as they are selected the next set of features are made visible in the same panel supporting the key approach. The result panel is simultaneously updated with the matched taxa.

In the next section we motivate investigating ontology technologies as a mechanism to support taxonomy and how an ontology-based approach inherently differs from legacy computer aided approaches. These differences are not necessarily obvious when a user is only confronted with the user interfaces of computer applications such as an existing taxonomic key as presented in Lucid. Similar to initatives such as DELTA and Xper, the taxonomic knowledge is captured in a standard language (e.g. OWL), but with the main difference that the ontology language is logic-based. These decidable logic-based languages (e.g. Description Logics) are considered to be significant technological breakthroughs in Artificial Intelligence (AI) [32, p. 149] and therefore offer several additional advantages such as the standardised knowledge-base representations in a ontology language, as well as formalised reasoning that support fundamental classification and logic-based reasoning. The architecture of the software application is thus not based on legacy and vendor specific formats, but uses internationally approved and generic standards such as OWL2 for the ontology language and the standard reasoning packages available based on these standards.

Male. Length 9.3–11.1 mm; integument usually black, T6–T7 orange in *M. seewaldi*; clypeus gently convex, short, clypeo-ocellus:clypeus 1:0.7; antennal flagellum convex ventrally; vertex short, orbit–occiput:orbit–ocellus 1:10; 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
- Female clypeus unmodified, ventral edge entire and horizontal. 4
2. Ventrolateral region of clypeus flat; vestiture of T1white. . . . . *Megachile trichroma* Friese.
- Ventrolateral region of clypeus tuberculate; vestiture on T1 black or white 3
3. Ventrolateral region of clypeus stongly tuberculate; vestiture on T1 black . . . . *Megachile cornigera* Friese
- Ventrolateral region of clypeus slightly convex; vestiture on T1 white . . . . *Megachile braunsiana* Friese
4. Metasoma clothed mostly with bright orange vestiture 5
- Metasomal vestiture partly white 6
5. Head and mesosoma brownish–orange. . . . . *Megachile dorsata* Smith
- Head and mesosoma mostly black . . . . *Megachile cognata* Smith
6. Metasomal dorsum with incomplete white distal fringes, black vestiture mesally, white hair semi–erect . . . . *Megachile ianthoptera* Smith
- Metasomal with complete pallid, appressed distal fringes 7
7. Mesosomal dorsum mostly with reddish–orange vestiture; male T6 acutely pointed posteromedially . . . .*Megachile discolor* Smith
- Mesososomal dorsum with whitish vestiture, if yellowish–brown metasomal distal fringes concolorous with mesosoma; male T6 concave posteromedially 8
8. Metasoma entirely clothed with white vestiture; large species, 25 mm long . . . . *Megachile rufoscopacea* Friese
- Metasomal terga black anteriorly, pallid posteriorly; small to medium sized bees, less than 19 mm long 9
9. Female T6 black; occurring in the South–western Cape Province . . . . *Megachile serrula* **sp. n.**
- Vestiture on female T6 at least partly white; occurring north of the Vaal River (South Africa). 10
10. Female T6 black, male T6 white. . . . . *Megachile angulata* Smith
- Female and male T6 yellowish to orange . . . . *Megachile seewaldi* Strand

Figure 1: An example of a taxonomic single access (dichotomous) key.

2.2. Taxonomy, Classification and Ontologies

Classification and categorization are fundamental human endeavors that are intuitively used by humans to make sense of the world [33]. How cognitive models are formed and how people use concepts and categories have developed into a branch of philosophy on classification [34], [35]. A key aspect of classification is the definition and description of categories or concepts based on shared characteristics, as well as the definition of a hierarchy of concepts, also called a *taxonomy*[†], related though an is-a relation [36, p. 3,4]. The definitions of concepts strongly depend on the context in which the specific definition makes sense. Examples include the definition of concepts such as *animal* and *mammal* within the context of the natural world (and not, for instance, to describe behavior) where an *animal* can be defined as *a vertebrate that is not an insect or a plant*. *Mammals* could then be defined through an *is-a* relation as a *mammal* is-a *animal that is warm blooded*.

Taxonomy within the context of the life sciences originated with the original work of Carl Linnaeus as published in his *Systema Naturae* [37]. The Linnaean taxonomy is a particular form of biological classification dividing all living organinisms into three kingdoms. These kingdoms are divided into *classes*, and they, in turn, into *orders*, *families*, *genera* (singular: *genus*), and *species* (singular: *species*) and an additional rank lower than species [38].

Computer programs have never been particularly effective at classification based on descriptions of qualitative characteristics, or semantics. In order to do classification, categories or classes are translated into digital models such as databases or the spreadsheet data format used by Lucid where rows are used to provide a textual description of a characteristic, and columns (indicative of a genus or species) would use a 1 or 0 to indicate the absence or presence of this characteristic. None of the qualitative semantics is represented. Rather, the semantics is programmatically interpreted and coded. Not only do any changes require code modifications, but the context and intent of the coding in this format often get lost if not properly documented. The richness and semantics or meaning of the qualitative descriptions such as, for example, in the case of Afrotropical bees, a characteristic stating '*Female clypeus distinctly modified, ventral edge concave mediolaterally, pointed or tuberculate medially*' are not captured at all.

The logic basis of computational ontologies provides several advantages for the computerized capturing and manipulation of qualitative knowledge and assertions such as the morphological descriptions for taxonomic knowledge bases. Furthermore, the model-theoretic semantics of OWL2 ontologies is set-based, which makes them particularly suitable for classification problems. Classification is the essence of most functions in biological taxonomy [39], [40]. To illustrate, it is possible to declare that a *genus* consists of *a set of bees* that all have a particular *set of characteristics*, which could be the set of characteristics the Afrotropical bee key uses to uniquely describe a particular taxon. The fact that biological taxonomy is in essence a classification of organisms based on qualitative characteristics, and that ontologies provide a means to capture such knowledge and are particularly suited for classification, provided one of the motivations for this study that investigated the use of computational ontologies to support taxonomic functions.

––––––––––

[†] In contrast with the term taxonomy as used in biological taxonomy, this taxonomy is just a hierarchical organization of concepts.

2.3. Ontologies and Ontology Technologies

In this section we provide a more in-depth introduction to ontologies and the technologies used within the project. Ontologies allow for a means to capture qualitative descriptions and do classifications, and we therefore explored how morphological key data such as discussed in the previous section could be represented in a formalized computational ontology.

Ontologies have roots in philosophy as indicated by the most common definitions for an ontology. According to Webster's Dictionary an ontology is a *branch of metaphysics relating to the nature and relations of being* or *a particular theory about the nature of being or the kinds of existence* [41]. Studer et al. defines an ontology as *a formal, explicit specification of a shared conceptualisation* [24]. However, the meaning of the term *ontology* evolved and it is used today for many related notions in various disciplines. Within artificial intelligence (AI) the term *ontology* has developed into a description of a knowledge representation artefact based on logic [42], [43]. Relative recent developments in computational logic, and specifically the development of description logics (DLs), has led to the notion of *computational* ontologies, which is a more specific application of the general notion of ontologies in AI [44], [45]. Computational ontologies provide a mechanism to capture qualitative descriptions and assertions about concepts and relations in a domain using a logic language, which in turn, allows standard reasoners to draw inferences from these descriptions [46]–[48]. A domain or field of knowledge can thus be represented using an ontology and an ontology within computer science is often defined as a shared, formal explicit specification of a conceptualization [49].

Computational ontologies are knowledge representation artefacts based on description logics (DLs), which are decidable fragments of first-order logic [44], [50]. The fact that DLs are decidable had a significant impact on the development of ontology languages for the construction of computational ontologies as computer readable knowledge bases. The specification of OWL, the Web Ontology Language, as a W3C Semantic Web standard for expressing meaning and semantics in ontologies supported this adoption. Standardization of such a language necessarily results in awareness and coordinated development [46]. OWL version 1.0 was originally published in 2004 [46]. OWL2 is an extension of the original version and OWL2 was first published in 2009, with a second edition published in 2012 [51], [52]. The specification of OWL2 resulted in the creation and publication of several OWL ontologies, reasoners that can infer consequences from OWL2 ontology assertions and several tools that support ontology editing and use, for instance, by using the standard OWL2 API [53]–[55].

Description Logics, Reasoning and Reasoners

DLs were adopted as the basis for OWL2 (the Web Ontology Language) by the W3C. Through the use of OWL, DLs are widely used in representing specific domains of interest in ontologies [46], [51], [56]. DLs are equipped with a precise model-theoretic semantics and have the capability of computing inferences, or in other words, providing reasoning over ontologies [50]. Because of the standardization of OWL2 by the W3C, a variety of tools and reasoners have been developed that support the construction, use and querying of ontologies. An example of such a tool is the ontology editor Protégé [57] that is distributed with packaged reasoners including Hermit [58], Pellet [59] and Fact++ [60], [61] to query and reason using OWL2 ontologies.

An ontology uses axioms or sets of syntactic sentences, also referred to as assertions, when referring to the knowledge representation aspect. These assertions represent the knowledge of a specific domain of interest,

and the assertions are composed of a set of concepts, roles and individuals[‡] [46]. Concepts are used to represent categories, roles link such categories or concepts and individuals belong to concepts and categories. Reasoning is a means of deriving implicit knowledge from a set of explicitly stated assertions in a logic-based ontology. Given that the packaged reasoners are implemented based on the OWL2 specifications and standards, these reasoners implement some standard reasoning services:

- *Concept Satisfiability*: Satisfiability testing is the process of checking and determining if a concept contains any individuals, i.e. if the concept is a valid concept given the domain representation. A concept is said to be unsatisfiable if it cannot have any individuals or instances in any model of the ontology.
- *Subsumption Checking*: Subsumption testing is when checks are done to ascertain whether a concept is a sub-concept or a super-concept of another. Subsumption testing is based on the 'is-a' role or is-a relation [62] that is widely used in ontology engineering to represent the *taxonomy* of an ontology or the hierarchy of concepts from general to specific.
- *Consistency Checking*: Consistency checking is the non-trivial reasoning task of ensuring that the set of assertions made in the ontology do not contain any explicit or implicit contradictions, or formally, whether an ontology has a model or not. If there is a model of the ontology, it means that there is an interpretation of it that satisfies all the sentences or assertions in the ontology. On the other hand, if no model exists then the ontology is said to be inconsistent.
- *Instance Checking*: Instance checking is the reasoning task that determines whether a specific individual within an ontology belongs to a concept or complex concept.

2.4. Related work: Ontologies within Bio-sciences

The standardization of terminology such as discussed in section 2.1 using basic semantic definitions and enabling semantic interoperability is one of the functions often supported by ontologies. In particular, ontologies are widely adopted within the bio-sciences [23], [63]. Examples of large projects that are based on ontologies include, but are not limited to, the list below:

- The Gene Ontology essentially describes gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. With new discoveries the ontologies are constantly being reviewed, and consists of in the region of 40,000 biological concepts [64]–[68].
- SNOMED CT is a large-scale ontology comprising clinical health information and it is the most comprehensive clinical vocabulary available. SNOMED CT is concept-oriented and uses an advanced structure that meets most accepted criteria for a well-formed, machine-readable terminology [69].
- BIOTOP is described as a top-domain ontology. Such an ontology is used to standardise terminology and the basic vocabulary in a domain by providing definitions for the foundational entities of the domain and to unambiguously describe facts in this domain [70]–[72].
- EnVO, the Environment Ontology, is an ontology of environmental features and habitats [73], [74].

─────────

[‡] The terminology for ontologies are often confusing. The terms classes, relations and instances are often used instead of concepts, roles and individuals respectively. For the purpose of this paper we adopt the DL terminology, which specifies formal semantics for the terms concepts, roles and individuals [45].

Several communities focusing more on systematics and taxonomy created and used ontologies, such as the NCBI Taxon Ontology [75] and the Vertebrate Taxonomy Ontology [76], [77]. Some of these ontologies are listed in the OBO Foundry [78] or BioPortal [79]. The focus of most of these ontologies is generally on interoperability (or allowing integration of knowledge often using some kind of meta-model), shared terminologies and standardized vocabulary, and meta-data annotations through the modeling of the various aspects of different species and taxa. Examples of these ontologies and approaches include:

- The work of Dahdul et al. [80] which reports on a study that uses the rich and formal semantics of ontologies to curate a collection of 47 phylogenetic studies on ostariophysan fishes and their relatives with the goal of integrating these complex phenotype descriptions with information from an existing model organism database of zebrafish;
- The work of Dececchi et al. [81] which uses ontologies and reasoning to synthesize morphological knowledge and phenotypic data across discrete publications in the Phenoscape Knowledgebase [82]. They present a methodology that use ontology-based reasoning systems to automatically integrate large amounts of evolutionary character state descriptions into a synthetic character matrix of neomorphic (presence/absence) data;
- The work of Blank et al. [83] which reports on the development of an ontology of microbiological terms, including prokaryotic qualities and processes, material entities (such as cell components), chemical entities (such as microbiological culture media and medium ingredients), and assays. This ontology called MicrO, has as purpose interoperability by connecting microbial classes with related conecpts such as chemical entities, material entities, biological processes, molecular functions, and qualities;
- The ambitious work of Mungali et al. [84] which aims to develop an integrated cross-species ontology called Uberon [85]. Uberon represents structures in a species-neutral way including extensive associations to existing species-centric anatomical ontologies, and thus allows integration of model organism and human data. Uberon intends to provide a bridge between anatomical structures in different taxa for cross-species inference;
- The work of Dahdul et al. [86] which reports on the development of a unified anatomy ontology of the vertebrate skeletal system (VSAO), specifically to support comparative morphology or compare phenotypes across taxa. The main stated purpose is to share data across databases and to unify semantically similar concepts. VSAO aims to provide a module of anatomical concepts for the vertebrate skeletal system which unifies the existing terminologies in multi-species and single-species anatomy ontologies.

Closely related to the representation of taxonomic data in an ontology as proposed by this study are initiatives such as:

- FLOPO, the Flora Phenotype Ontology [87], [88], [88]. FLOPO mostly focuses on supporting the analysis of organismic traits specificallt on the plant community, by establishing a standardized vocabulary describing traits of plant species found in Flora;
- HPO, the Human Phenotype Ontology [89] aims to establish a standardized, controlled vocabulary that allows human phenotype information to be described in an unambiguous fashion in medical publications and databases. Such a standard ontology that capture phenotypic information allows for the use of computational algorithms that exploit semantic similarity between related phenotypic abnormalities to define phenotypic similarity metrics, which can be used to perform database searches for clinical diagnostics or as a basis for incorporating the human phenome into large-scale

computational analysis of gene expression patterns and other cellular phenomena associated with human disease [90]–[93];

- The Hymenoptera Anatomy Ontology (HAO) [94]. HAO focuses on the morphological modeling of *Hymenoptera*, such as sawflies, wasps, bees and ants. The HAO was developed to create a common and shared terminology for Hymenoptera Anatomy, as well as to provide a platform to integrate the corpus of information about hymenopteran phenotypes that is inaccessible due to language discrepancies [94]–[98];

- The Spider Ontology, which is part of the Tree Of Life: Phylogeny of Spiders initiative [99], [100]. The Spider ontology models spider comparative biology including anatomical parts (e.g. leg, claw), behavior (e.g. courtship, combing) and products (e.g. silk, web, borrow). Even though this ontology models some morhological aspects, it has, similar to the previous mentioned ontologies, as purpose interoperability and integration;

- TrOn, the Tribolium Ontology, is an anatomical or morphological ontology for the Beetle *Tribolium castaneum* [101]. TrOn describes morphological structures and how these structures relate to each other of the red flour beetle *Tribolium castaneum* (*Coleoptera*). TrOn mostly comprise of external morphological structures that are consistently annotated for the developmental stages larva, pupa and adult.

In this paper we propose that the taxonomic knowledge is captured as an ontology language (OWL), but also using constructs that support reasoning with the specific goal namely to assist with taxonomic functions. Ontologies are therefore not just a standardized vocabulary, or a mechanism to augment taxonomic knowledge as is done in most of the previously listed examples.

3. Experimental Case: Taxonomy of Afrotropical Bees

The research project was concerned with the specific case of the taxonomic revisions required for Afrotropical bees given collected specimens and data. Afrotropical bees are one of the largest pollinator groups in Southern Africa, which makes them a biologically important group of organisms [7], [26]. These bees use pollen and nectar that have been collected from flowers to feed their progeny. When bees visit the flowers and crops to collect such products they end up pollinating them, which enables new seed production of many different species of flowering plants [102]. In additional to often-threatened wild flora, many agricultural crops are dependent on such bees for pollination in order to produce fruit and/or seeds. The *Catalogue of Afrotropical Bees* (*Hymenoptera: Apoidea: Apiformes*) [7] and the *Bee Genera and Subgenera of Sub-Saharan Africa* [26] both provide extensive taxonomic information on the valid names, nomenclatorial history of, and published references to the known bees of Sub-Saharan Africa and the Western Indian Ocean Islands (excluding the honey bee (*Apis mellifera Linnaeus*)). The catalogue adheres to the relevant taxonomic standards and thus contains the species with references as well as taxonomic changes such as new name combinations, with correct latinisation and gender. The catalogue also provides the distribution of species by country, plants visited, hosts (for parasitic bee species) and parasites, as well as the type's gender, depository and country locality for each of the described species [7].

Given recently collected data, taxonomists are revising the existing taxonomic knowledge base of Afrotropical bees, for instance as contained in the *Catalogue of Afrotropical Bees* (*Hymenoptera: Apoidea: Apiformes*) [7]. This revision work includes carefully analyzing and comparing existing descriptions and

groups to determine whether the new data refines or extends taxa. Typically this work would entail revisiting all taxa containing specific morphological descriptions and a tool that could, for instance, extract this knowledge from the taxonomic descriptions would aid the researchers substantially.

The first step in the project investigation was therefore to analyze the existing taxonomic knowledge base for Afrotropical bees. In *The Bee Genera and Subgenera of Sub-Saharan Africa* [26] a popular representation of the information pertaining to Afrotropical bees, biodiversity conservation as well as a taxonomic key for the bee genera and subgenera that occur in sub-Saharan Africa are presented.

There is a basic Lucid key available for Afrotropical bees[§] and the initial taxonomic knowledge we used in our experiment was originally formatted for Lucid because this was already a computer file and not text data. In order to create such a Lucid key, data from the keys in the catalogue and booklet of Afrotropical bees were used to develop an Excel-spreadsheet in a standard format prescribed for Lucid that is available on the Web. Such a spreadsheet consists of two main areas namely the actual taxa (in this case the bee genera (or species)) and the key features that describe the specific taxa (or the bee genera). Using the example of the Afrotropical bees, a specific bee genus is uniquely described or identified by a set of morphological features. For example, the bee genus *Afrodasypoda female (Melittidae)* is associated with the following features:

- `General:Gender (female/male): Female,`
- `Head: 12 antennal segments,`
- `Metasoma: Six metasomal terga,`
- `General: Scopa (pollen basket) position: Hind leg, not corbicula.`

4. Experimental Development

In an experimental study the first task to be executed is the elicitation of the application requirements. In this case we interviewed taxonomists, specifically an Afrotropical bee taxonomist who was interested in exploring the use of ontology to taxonomy. We identified several requirements listed below:

- The first requirement identified was to capture the taxonomic knowledge in an ontology in such a way that the domain expert could relate. The ontology should therefore be a reusable knowledge base for the taxonomic knowledge of Afrotropical bees in addition to the existing textual descriptions. It should also be possible to update and maintain the ontology as is necessary, for instance after new data or taxonomic revision results. The starting point for this work was the excel spreadsheets of morphological characteristics used to describe taxa that were prepared for the Lucid Key for Afrotropical bees referred to in Section 2.1.
- The second requirement was to implement identification functionality or a *key* to the Afrotropical bees similar to that of Lucid but using an ontology-driven architecture with standard reasoning services.

[§] http://africanpollination.org/Africanbeegenera/Key_to_african_bee_genera.html,
  Need a browser that will run Java with the correct Java installation and security preferences for Lucid.

- The third requirement was to support taxonomists with revision work, i.e. when scrutinizing a bee taxon, typically a taxonomist would like quick access to all possible taxa that exhibit similar morphological characteristics and this functionality should be possible using the developed ontology and available reasoners.
- Finally, since taxonomists do not require ontology engineering skills and using Protégé, it was identified as a requirement that it is necessary to develop a web-based and user-friendly front-end to the application that supports the identified taxonomic functions (namely identification and revision).

These requirements were used to direct the activities of the experimental study as is discussed in the following sections. Section 4.1 presents the ontology, Section 4.2 discuss the reasoning support firstly with regards to identification, and secondly with regards to revision. Section 4.3 presents the application development of the WOC, the Web Ontology Classifier.

4.1. The Afrotropical Bee Ontology

The first and core application component that had to be developed in the study is the ontology[**] that captures the Afrotropical bee taxonomic knowledge as discussed in the previous section and published in previous work [103]. The initial intent of the ontology development focused on capturing the taxonomic knowledge of Afrotropical bees in such a way that it supports the identification function similar to a taxonomic key. We therefore used the Excel-spreadsheet that was developed as input into the Lucid key as basis (see Section 2.1) and adopted Protégé as ontology editor. The two core concepts that have to be represented in the ontology are the *taxon* (in this case Afrotropical bee genera) and the morphological characteristics or *diagnostic features*. A diagnostic feature in the ontology consists of some body part of a bee with a description, e.g. *General:Scopa (pollen basket) position:Hind leg, not corbicula*. We made the modeling decision to represent a diagnostic feature as a *BodyPart* that has a feature (such as colour or shape). A bee body consists of body parts that may consist of other body parts and so on until we reach the body part that has a specific morphological feature that can be used to identify the bee such as the *'colour of the integument that is part of the thorax'* or the *'shape of a segment that is part of the head'*.

The main concepts (or classes given Protégé terminology) in the ontology are therefore:
- The `BodyPart` hierarchy that has sub-concepts such as Appendage that can be an Antenna, Leg, Scopa etc. given the morphological description, or Regions such as Clypeus, or Integument.
- The concept hierarchies necessary to describe the morphology such as Colour, Structure, Position etc. Refer to Figure 2.
- The Object Properties (or Roles) are mainly `hasPart` with sub-role `hasBodyPart`, `hasFeature` with sub-roles roles such as `hasColour`, `hasLength,` `hasPosition` etc, and `hasDiagnosticFeature`.
- The `DiagnosticFeature`: the concepts representing of morphological descriptions such as
    o `OcelliPositionDF ⊑ (Ocellus and hasPosition some InFrontOfVertexPosition)`

─────────

[**] The ontology can be found on github7 - https://github.com/Nish01/OntologyClassifierApplication.

      o   `PropodiumIntegumentColourBlackDF ⊑ (Propodium and (hasPart some`
          `(Integument and (hasColour some BlackIntegumentColour)))`

- `Taxon`, the taxa that are linked to the sets of DiagnosticFeatures. For the implementation of the key for the genera of Afrotropical bees, a specific bee genus would be uniquely identified by sets of diagnostic features. A specific genus, *Afrodasypoda* and a few of its diagnostic features in Protégé are depicted in Figure 3.

Figure 2: The Afrotropical Bee Ontology in Protégé

Throughout the development in Protégé, standard packaged reasoners were used to test and query the ontology and the domain expert, the bee taxonomist, was involved. A pragmatic ontology engineering approach was adopted to keep the descriptions and modeling as close as possible to the existing domain data's structure to allow the domain expert to participate in the knowledge capturing. In order to provide some insight into the size of the ontology and the DL expressivity, some of the Protégé metrics of the ontology are: Axioms: 10426, Logical axiom count: 7129, Class count: 1182, Object property count: 21 and DL Expressivity: $\mathcal{ALCHQ}$. We found that Protégé provided all the necessary functionality for the development of the ontology and although the ontology has 10426 axioms, we did not encounter storage or performance problems. The main challenges experienced during this phase included the translation of the domain or taxonomic knowledge into an OWL2 ontology as well as the communication with the domain expert to explain what we did and for clarification of vague aspects. Logical assertions and ontology models are conceptual and the translation of domain knowledge into such a model is often not trivial, for example translating the sentence 'Mesosoma brownish-orange' into the ontology axiom *Mesosoma is a bodypart that hasDiagnosticFeature a Feature that is a Colour which is Brownish-orange* required extensive explanations.

4.2. Reasoning Support

As stated we define *taxonomy* as the scientific classification, description and grouping of biological organisms into hierarchies based on sets of shared characteristics, as well as documenting the principles that enforce such classification. We identified 2 main types or categories of *taxonomic functions* namely the classification functions concerned with *identification* and naming of organisms, as well as classification

functions concerned with categorization and *revision* (i.e. grouping and describing, or revisiting existing groups and descriptions). If the taxonomic descriptions already exist such as in the case of Afrotropical bees, the second category of taxonomic fuctions will primarily focus on revision. Given the requirements as discussed in the beginning of Section 4, the second task was to implement *identification* over the ontology given standard reasoning support.
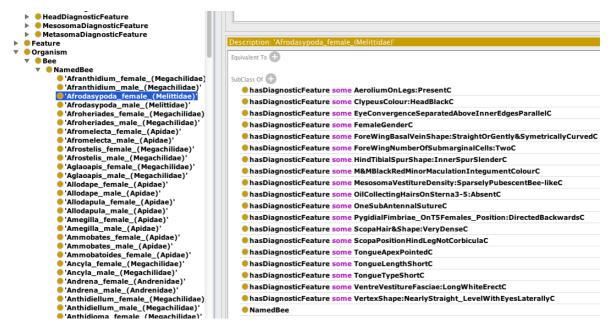

Figure 3: The genus 'Afrodasypoda' with its associated diagnostic features

Taxonomic Identification:

Identification or, in other words, the functionality presented by a taxonomic key is implemented given the queries to the reasoner over the Afrotropical bee ontology as below. The queries are first presented in natural language and then translated into DL syntax. Referring to the Afrotropical bee ontology, a particular bee genus (or species) has a set of diagnostic features and the collective set of these features uniquely identify the particular genus or species. In DL syntax the assertion or axiom is:

Axiom 4.1: (Bee Taxon and Diagnostic Features)
```
BeeTaxon ⊑ ∃ hasDiagnosticFeature.DiagnosticFeature
```

Using an actual example for identification using Axiom 4.1:
```
Plesianthidium (Spinanthidiellum) volkmanni (Friese) male ⊑ ∃
hasDiagnosticFeature.MaleBasitarsusColour:Black
```
where '*Plesianthidium (Spinanthidiellum) volkmanni (Friese) male*' is the genus of Afrotropical bee that has a basitarsus (the first tarsal segment in the leg of an insect) that is

black, modeled as a diagnostic feature 'MaleBasitarsusColour:Black' in the ontology.

In terms of classification, when attempting to identify a bee, a user will examine and analyze the specimen and identify the distinct characteristics such as a black basitarsus. These characteristics will then be selected in the application and the result should result in the identification of the appropriate taxon (genus or species). This function can be translated into a query (defined as Query 4.1) namely '*which genera or species (or objects, in a general situation) exist that have these selected set of diagnostics features?*' Translated as a DL query asked over Axiom 4.1 to retrieve the desired results:

Query 4.1: (Find Taxa Query):

$$\exists\ \texttt{hasDiagnosticFeature.DiagnosticFeature} \sqsupseteq \texttt{BeeTaxon}$$

the unknown to be retrieved is the `DiagnosticFeature`, which is a *filter*[††] for the existential restriction ($\exists$) *or* which taxa `hasDiagnosticFeature some DiagnosticFeature?`, where `DiagnosticFeature` is the feature selected by the user.

The reasoner would then return all the taxa that are related to the set of selected diagnostic feature(s) via the role `hasDiagnosticFeature`, which are the genus or species that have those selected characteristics as depicted in Figure 4 in Protégé. The results may be a number of taxa depending on the uniqueness, and the number of features used, but as more are found and added to the query, the list of taxa becomes more refined. In order to execute these queries, the standard subsumption checking reasoning services could be applied.

Figure 4: Identification Query in Protégé

[††] The term filler is standard DL terminology for the quantifiers of existential axioms. In this instance the `DiagnosticFeature` is termed the filler of an existential query. Standard reasoning services as packaged in Protégé do not allow for queries using the fillers of existential axioms but some extensions to standard reasoning were developed [104].

We evaluated the application against the Lucid key, and we obtained exactly the same results with regards to identification of a taxon or taxa. The ontology and standard reasoning services available thus support identification taxonomic functions. This functionality could be incorporated into an application as a multi-access key that allows the user to identify relevant taxa.

Taxonomic Revision:

As stated, the second category of taxonomic functions identified is functions concerned with categorization and *revision* (i.e. grouping and describing, or revisiting existing groups and descriptions). If the taxonomic descriptions already exist such as in the case of Afrotropical bees, these taxonomic fuctions will primarily focus on revision i.e. revising Afrotropical bee genera or species based on a combination of morphological characteristics. This use case scenario implied that it is necessary to pose a query that is the opposite of the identification query, or the reverse of Query 4.1 namely '*what diagnostic features does a selected taxon (genus or species) exhibit?*' Since ontologies are set based, this query is necessarily expanded to use a *set of taxa* that could be selected, which results in the query '*what set of diagnostic features does a set of selected taxa exhibit and in this set, which of the diagnostic features are common to all the taxa in the set of taxa?*'

Referring to Axiom 4.1 from the Afrotropical bee ontology:

```
BeeTaxon ⊑ ∃ hasDiagnosticFeature.DiagnosticFeature
```
or
```
∃ hasDiagnosticFeature.DiagnosticFeature ⊒ BeeTaxon
```
the unknown to be retrieved is the `DiagnosticFeature`, which is a *filler*[‡‡] for the existential restriction (∃)

When we explored the reasoning support for these type of queries, it became clear that standard reasoners packaged with the Protégé standard installation (such as Fact++, Pellet and Hermit) do not support these types of queries.

We explored several mechanisms to allow for these queries over the bee ontology:

- Assert inverse axioms: This solution would entail asserting the axioms in the ontology that state the inverse assertions using an inverse object property, For each `BeeTaxon` `hasDiagnosticsFeature some DiagnosticFeature` we could assert `DiagnosticsFeature isDiagnosticFeatureOf some BeeTaxon`. Using standard reasoning as packaged with Protégé, this solution did return the correct results. This solution does have semantic implications because it asserts that each `DiagnosticFeature` must have at least one relation with a `Taxon` via the `isDiagnosticFeatureOf` role, which is a stricter modeling assertion that just stating that each `Taxon` has a diagnostic feature. One reason why we decided against this solution was pragmatic since one of the requirements we identified for the project was that the bee ontology should function as a reusable knowledge base that models the

---

[‡‡] The term filler is standard DL terminology for the quantifiers of existential axioms. In this instance the `DiagnosticFeature` is termed the filler of an existential query. Standard reasoning services as packaged in Protégé do not allow for queries using the fillers of existential axioms but some extensions to standard reasoning were developed [104].

taxonomic knowledge of Aftrotropical bees. The ontology would require additional assertions for each bee taxon and this approach could lead to maintenance problems if future ontology revisions add assertions in the ontology without adding the necessary inverse axioms.

- Remodeling using individuals or ABox assertions: Another solution is to use individuals for the bee taxa and diagnostic features rather than concepts in the bee ontology but this solution required a complete remodeling of the ontology and knowledge base. In addition, reasoning using individuals is generally less effective, especially when dealing with large ontologies.

Because the ontology captures the taxonomic knowledge for Afrotropical Bees in a standardized language (OWL2), the ontology as an artifact and contribution should be reusable and maintainable, and any modeling choices should support these requirements. Therefore, rather than opting for remodeling of the ontology, we explored extended reasoning services, which means that the standard reasoning over an OWL2 ontology is extended with additional algorithms. A reasoning algorithm that allows for querying fillers existed as a plugin for Protégé [104]. As stated, a standard DL query (such as Query 4.1) can only compute the subclasses or superclasses of the query class expression through the subsumption checking reasoning service. The existential query plugin use standard reasoners such as Pellet or FaCT++ but add another algorithm to compute the existential filler query.

With reference to Axiom 4.1:
$$\text{BeeSpecies} \sqsubseteq \exists \ \text{hasDiagnosticFeature.DiagnosticFeature,}$$
the (implied) filler would be the DiagnosticFeature, which is what is required by Query 4.2.

The algorithm essentially performs a series of tests using the concept hierarchy in the ontology to optimize the process. The first step is to compute the complete concept hierarchy through the standard subsumption testing reasoning service. The extended reasoning algorithm then use each of the concepts and checks to see if this concept is a filler for the role and returns a set of the fillers. Formally, given an initial base class C and a property R the algorithm then checks to see if $C \sqsubseteq \exists \ R.\text{Thing}$ is entailed. If it is entailed, iterations are made over the subclass of Thing. For each subclass, Cn, a check is then done to see if $C \sqsubseteq \exists \ R.\text{Cn}$ is entailed. If it is entailed the process is repeated with the subclasses of Cn. Once all the entailments have been checked, the fillers in the restrictions of entailed axioms (Cn), are collected and returned.

Figure 5 : The Existential Query Plugin in Protégé.

4.3. Application Development

The individual components, namely the ontology and the reasoners discussed previously, are standardized and reusable components, but given the requirements of taxonomists, the components were integrated into an application. An initial version of the application that implemented the identification function (or the ontology-driven key) was a standalone application. However, portability and accessibility were challenges as the application had to be installed on each required platform and the latest version of the ontology had to be distributed to each installation. It was therefore decided to implement the final application as a web-based application.

The *web ontology classifier* (WOC) uses the developed Afrotropical bee ontology to support the taxonomic functions *identification* and *revision* given the taxonomy of Afrotropical bees. Using the WOC a user makes a selection of a set of diagnostic features and the application would then display the bee taxa that are associated with the selected features. In addition, to assist with taxonomic revisions the user could select a set of bee taxa, and the application would display the corresponding *common* (intersection), *uncommon* (union) and *remaining* diagnostic features of the selected taxa. The Web Ontology Classifier (WOC) Application was developed using Play Frameworks and Java[§§].

The WOC launches with a home page with basic instructions on how to use the application. By clicking the menu bar link, '*Classify and Query*,' the user will be directed to an upload page where an ontology can be selected and uploaded. Upon submission of the ontology, the reasoning functionality necessary to support the taxonomic functions over the ontology are loaded and the ontology is classified. Once classified the ontology will be displayed on the page shown in Figure 6. 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. On the far right is a key showing what each colour represents. The user has two options:
  1) Select a set of diagnostic features from the first column and submit (to do an identification query); or,
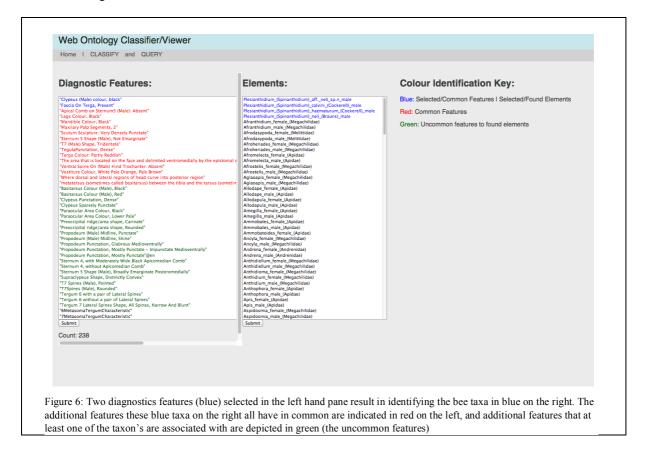 2) Select a set of bee taxa on the right and submit (doing a query that supports revision).

Figure 6 depicts the interface after the user originally selected two diagnostic features (shown in blue at the top of the left column), and four bee taxa elements were found that satisfied the query (also shown in blue at the top of the right column). In addition, the existential query is automatically executed updating the left hand pane, and the elements listed in red in the first column are the diagnostic features that are *common* to all the bee taxa. The green features 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, or, in other words, the diagnostic features that are held by at least one of the bee taxa but not shared by the whole group of identified taxa. The features in black are the *remainder* of the 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 source code as well as the ontology and instructions to setup the application offline can be found on github7 - https://github.com/Nish01/OntologyClassifierApplication.
  • An online hosting is available at http://41.185.28.217:9000/

For evaluation purposes the WOC application was used by the original domain expert and refinement of the user-interface was done given his feedback. However, in addition, we presented the WOC to a group of approximately hundred biologists from twenty different South African institutions during a focus group workshop where the research project and application was introduced and demonstrated. We collected feedback with a survey and received 15 completed survey forms back (all members in our audience were not taxonomists or interested in such an application as the WOC). In addition, there was a lot of discussion on the survey questions and several verbal discussions were recorded and included in the feedback consolidation. The taxonomists that gave verbal feedback were mostly unsure about the technological aspects and felt verbal feedback was more appropriate. All the taxonomists that participated in the survey or gave feedback otherwise were supportive of the research project and the application, and voiced the opinion that such a tool will assist them in their taxonomic work. Many of the taxonomists were interested in adopting the WOC. The fact that the ontology is a standardized and reusable knowledge base of the taxonomic knowledge was identified as an advantage, although how such an artifact will be integrated in the standard revision practices was not clear. A few key areas for improvement were also identified such as the user friendliness of the interface, the provision for more features such as pictures of the species or features, and finally, the ability to extend and alter the taxonomic ontology with a dedicated interface and thus not being dependent on an ontology engineer or a tool such as Protégé.

Figure 6: 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)

5. Findings and Discussion

We identified the need for computerized support for taxonomy given the taxonomic impediment, in the case of this project, the taxonomic impediment and how it related to Afrotropical bees. Afrotropical bees is one of the most significant pollinator groups within the biodiversity of South Africa and only a few taxonomists are still actively focusing on this group, most of them on the verge of retirement. We therefore initiated the project to investigate whether ontology technologies could support taxonomic functions, which we grouped into two main categories namely identification and revision.

We identified four main requirements, which also translated into the main tasks of the project. The first was to capture the taxonomic knowledge base working with the bee taxonomists in an ontology. Secondly we identified reasoning support over the ontology for identification, and thirdly for revision. Finally we integrated the components and implemented web-based user-interface in order to develop the Web Ontology Classifier (or WOC) standalone application. All the requirements were met, namely:

1. The taxonomic knowledge base was represented in an ontology. The ontology development was done in conjunction with the Afrotropical bee taxonomist.
2. Identification was supported over the ontology used the standard reasoning services such as packaged with Protégé. The core construct we used in the ontology allows for the support of identification taxonomic function with available standard reasoning packages.
3. Revision was supported over the ontology but with extended reasoning services.
4. It was possible to integrate the components into a standalone application with a dedicated user interface, namely the WOC.

We evaluated the final WOC also against the Lucid key implementation and the results were similar with regards to the key or identification services, albeit with somewhat enhanced performance. However, the WOC application did not have the additional graphical support such as pictures of the bees that were available in the Lucid key. In addition, the domain expert as well as a group of taxonomists evaluated the application after a focus group discussion and the feedback was overall positive.

One of the advantages of the WOC is that it eliminates the use of Protégé, which could be a challenge for users not familiar with ontology engineering. However, it is possible to edit or augment the ontology whenever needed as well as check the reasoning services using Protégé without impacting the WOC. Once a new version of the ontology has been constructed, it is loaded into the WOC and the functionality remains. Similarly, a new version of any of the reasoning services could be tested in the Protégé environment and seamlessly used within the WOC.

The core modeling construct we used in the ontology is given in Axiom 4.1 namely `BeeTaxon` $\sqsubseteq$ $\exists$ `hasDiagnosticFeature.DiagnosticFeature`. The core concepts can be extended into the necessary hierarchy, e.g. the *DiagnosticFeature* is a complex concept namely a *BodyPart* with specific characteristics or features (such as colour or shape) as presented in Section 4.1. A body part may consist of other body parts and so on until we reach the body part that has a specific morphological feature that can be used to identify the bee such as the '*colour of the integument that is part of the thorax*' or the '*shape of a segment that is part of the head*'. One of the advantages of the core-modeling construct is extensibility in that all characteristics that can be used for any taxonomic identification could be modeled, no matter the family. It is possible to model that a Red-bearded bee-eater (*Nyctyornis amictus)* hasDiagnosticFeature red feathers on

neck (where the body part would be a neck that has a cover feathers which are red). Futhermore, it is possible to extend the *DiagnosticFeature* to not just be a *BodyPart* (e.g. morphological characteristic) but also other features used for identification such as area, diet or behavior. A *DiagnosticFeature* could therefore be an area in which the species breed.

The ontology model and core-modeling construct allowed for the support of the identified taxonomic functions with available standard and extended open-source reasoning packages. We consider this as an important finding since it was not necessary to do specific program development to support the core functionality. The main development efforts were in ontology engineering and the customized Web-based front-end. Furthermore, the core-modeling construct we used in the ontology development provides a first version of a design pattern for representing taxonomic information (that is, for example, the use of diagnostic features). This pattern is specifically focused on harnessing the associated advantages of computational ontologies namely the reasoning support.

With regards to the confidence of the results, two aspects require consideration: the first aspect relates to the correctness of the results provided by WOC. WOC is based on reasoners that have been proved to be correct. This means that the reasoning mechanism itself is correct. As with any software application, there may be implementation errors, though, which are addressed through a software testing activity. The second aspect relates to a chosen taxon being correct. In this case, correctness is dependent on the information provided by the user of the system. We tested the identification functionality against the existing Lucid key application and the results of the WOC were similar. Accuracy in this case is thus limited by the accuracy of the taxonomic knowledge captured in the ontology. As long as the relevant diagnostic features are linked to a taxon via the core ontology construct, the correct results will be returned by the reasoning services.

With regards to existing computerized keys, our application compares with regards to identification functionality once the taxonomic knowledge has been captured in the ontology, but with enhanced performance because once the ontology has been loaded in the reasoner or *classified*, queries have limited computational overhead.

One of the most significant advantages of the WOC architecture with the three components namely the ontology, the standard reasoners and the interface as discrete components is that any taxonomic ontology can be uploaded, provided that it is structured according to the requirements of WOC. It is, therefore, not completely all-encompassing, but nevertheless considerably more flexible than other available systems.

Existing key applications often have additional user interface support such as images of specimens. In the WOC our intended audience was taxonomists who do not need extensive user support, however, this was identified by the focus group evaluation as a further requirement. The WOC does not currently support such extensions except if built into the user interface, which has the disadvantage that it will require a programmatical update of the interface if the ontology is updated. Future research could investigate how to enhance the core model or application architecture with additional information such as images to ensure that this information is part of the taxonomic knowledge models.

The results of our work support standardization efforts within taxonomy such as Xper and DELTA, as well as the general vision TDWG. Ontologies represented using OWL adhere to a standard semantic language of

the W3C. Further work would entail investigating how the core-modeling construct of our model could be integrated into existing standards but still support the reasoning advantages. OWL is based on RDF, which follows the basic triplet semantic model namely object-predicate-subject, and this model has been shown to be a basis of semantic standardization [105]. Given the results of our study, future work could investigate how standard reasoning services could support existing standards.

With regards to the ontology development, further work would include a dedicated interface allowing taxonomists to extend their specific ontology but restricted to the core-modeling construct that is supported by the reasoning. It should be possible for a taxonomist to link taxa to their identified diagnostic features and thus extend the ontology.

Future work also include the refinement of the ontology model given ontology engineering approaches, as well as existing upper and domain ontologies. As stated the core modeling construct we used in the ontology is given in Axiom 4.1 namely `BeeTaxon ⊑ ∃ hasDiagnosticFeature.DiagnosticFeature`. The linking of these constructs to standard upper ontologies would refine our ontology. In addition, the existing modeling of diagnostic features could be refined using existing taxonomic ontologies such as HAO. However, as is the nature of all semantic collaborative efforts, such as effort would entail a working group with taxonomists to ensure that definitions of the diagnostic features conform to all ontologies.

6. Conclusion

Using the specific case of Afrotropical bees, this research study explored 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) exploring and extending the use of available reasoning algorithms to draw inferences that support the necessary taxonomy workflows, and 3) the implementation of a Web-based application, the WOC, integrating the ontology and reasoning services for taxonomy users. The WOC application supports the two broad categories of taxonomic functions we identified namely identification (keys) and taxonomic revision. The WOC support the identified requirements for an ontology-based system to support taxonomists, and futhermore, received a positive evaluation from a taxonomists in a workgroup.

When considering the taxonomic impediment, especially with regards to Afrotropical bees within South Africa, the results of our investigation are positive given the key requirements, and future work might assist with the global computerized support for taxonomy, which we believe is crucial given the sustainability of life on earth.

References

[1] CHM, 'Convention on Biological Diversity: Biodiversity Threats', *Convention on Biological Diversity*. [Online]. Available: http://www.biodiv.be/biodiversity/threats. [Accessed: 01-Oct-2016].

[2] Ronald E. Hester and Roy M. Harrison, *Biodiversity under threat.*, vol. 25. Royal Society of Chemistry, 2007.

[3] N. E. Stork, 'How many species are there?', *Biodiversity & Conservation*, vol. 2, no. 3, pp. 215–232, 1993.

[4] C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm, 'How Many Species Are There on Earth and in the Ocean?', *PLoS Biol*, vol. 9, no. 8, 2011.

[5] UNESCO, 'Education for sustainable development information brief: Biodiversity'. .

[6] J. M. Guerra-García, F. Espinosa, and J. C. García-Gómez, 'Trends in Taxonomy today: an overview about the main topics in Taxonomy', *Zoologica baetica*, vol. 19, pp. 15–49, 2008.

[7] C. Eardley and R. Urban, *Catalogue of Afrotropical bees (Hymenoptera: Apoidea: Apiformes)*. 2010.

[8] K. Elaine Hoagland, 'The taxonomic impediment and the convention on biodiversity', *Association of Systematics Collections Newsletter*, vol. 24, no. 5, pp. 61–62, 1996.

[9] D. E. Walter and S. Winterton, 'Keys and the Crisis in Taxonomy: Extinction or Reinvention?', *Annual Review of Entomology*, vol. 52, no. 1, pp. 193–208, Jan. 2007.

[10] Q. D. Wheeler, P. H. Raven, and E. O. Wilson, 'Taxonomy: impediment or expedient?', *Science (New York, NY)*, vol. 303, no. 5656, p. 285, 2004.

[11] Paolo Viscardi, 'Natural history collections –- why are they relevant?', *The Guardian*, 2016.

[12] H. C. J. Godfray, 'Challenges for taxonomy', *Nature*, vol. 417, no. 6884, pp. 17–19, 2002.

[13] H. C. J. Godfray, B.R.Clark, I. J. Kitching, S. J. Mayo, and M. J. Scoble, 'The web and the structure of taxonomy.', *Systematic biology*, vol. 56, no. 6, pp. 943–55, 2007.

[14] N. Maxted, 'Towards defining a taxonomic revision methodology', *Taxon*, pp. 653–660, 1992.

[15] M. R. de Carvalho *et al.*, 'Taxonomic Impediment or Impediment to Taxonomy? A Commentary on Systematics and the Cybertaxonomic-Automation Paradigm', *Evolutionary Biology*, vol. 34, no. 3–4, pp. 140–143, Dec. 2007.

[16] G. H. Dar, A. A. Khuroo, C. S. Reddy, and A. H. Malik, 'Impediment to Taxonomy and Its Impact on Biodiversity Science: An Indian Perspective', *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, vol. 82, no. 2, pp. 235–240, Jun. 2012.

[17] R. C. Nickerson, U. Varshney, and J. Muntermann, 'A method for taxonomy development and its application in information systems', *European Journal of Information Systems*, vol. 22, no. 3, pp. 336–359, 2013.

[18] N. MacLeod, Ed., *Automated taxon identification in systematics: theory, approaches and applications*. Boca Raton: CRC Press, Taylor and Francis Group, 2007.

[19] M. Sigovini, E. Keppel, and D. Tagliapietra, 'Open Nomenclature in the biodiversity era', *Methids in Ecology and Evolution*, vol. 7, no. 10, 2016.

[20] Peter Bengtson, 'Open Nomenclature', *Paleontology*, vol. 31, no. 1, pp. 223–227, 1988.

[21] TDWG, 'TDWG - Biodiversity Information Standards.' [Online]. Available: http://www.tdwg.org/. [Accessed: 11-May-2017].

[22] Deans AR, Lewis SE, Huala E, Anzaldo SS, and Ashburner M, 'Finding Our Way through Phenotypes', *PLOS Biology*, vol. 13, no. 1, 2015.

[23] P. N. Robinson and C. Webber, 'Phenotype ontologies and cross-species analysis for translational research', *PLoS Genet*, vol. 10, no. 4, p. e1004268, 2014.

[24] R. Studer, V. R. Benjamins, and D. Fensel, 'Knowledge engineering: principles and methods', *Data & knowledge engineering*, vol. 25, no. 1–2, pp. 161–197, 1998.

[25] S. G. Saupe, 'Taxonomic Keys', *Taxonomic Keys*, 2001. [Online]. Available:

http://www.uprm.edu/biology/profs/betancourtc/taxonomia3.htm. [Accessed: 11-May-2017].

[26] C. Eardley, M. Kuhlmann, and A. Pauly, 'The Bee Genera and Subgenera of sub-Saharan Africa', *ABC Taxa.*, no. 7, 2010.

[27] CSIROpedia, 'DELTA taxonomic computer programs', 2011. [Online]. Available: https://web.archive.org/web/20150320192646/http://csiropedia.csiro.au/display/CSIROpedia/DELTA+taxonomic+computer+programs. [Accessed: 03-Mar-2017].

[28] M. J. Cavalcanti, 'Free DELTA', *Free DELTA*, 2000. [Online]. Available: http://freedelta.sourceforge.net/. [Accessed: 12-May-2017].

[29] LIS, 'Xper2', *Xper*, 2017. [Online]. Available: http://www.infosyslab.fr/?q=en/resources/software/xper2. [Accessed: 12-May-2017].

[30] G. Norton, D. J. Patterson, and M. Schneider, 'LucID: A multimedia educational tool for identification and diagnostics', *International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International)*, vol. 4, no. 1, 2012.

[31] 'Lucid', *LucidCentral.org*. [Online]. Available: http://www.lucidcentral.com.

[32] Gautam Shroff, *The Intelligent Web*. Oxford University Press, 2013.

[33] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind.* Chicago and London: University of Chicago Press, 1987.

[34] A. J. H. Simons, 'Theory of Classification', *University of Sheffield: Anthony J H Simons*, 2006. [Online]. Available: http://staffwww.dcs.shef.ac.uk/people/A.Simons/classify/.

[35] Daniel Parrochia, 'Classification - Internet Encyclopedia of Philosophy.pdf'. 2007.

[36] K. D. Bailey, *Typologies and taxonomies: an introduction to classification techniques.* Thousand Oaks, California: Sage Publications, 1994.

[37] C. Linnaeus, *Systema naturae. 10th ed.*, vol. v.1. Laurentii Salvii, Stockholm, 1758.

[38] M. Taylor, 'What are classification, taxonomy, phylogeny, systematics and cladistics?', 2002. [Online]. Available: http://www.miketaylor.org.uk/dino/faq/s-class/terms/. [Accessed: 15-Jul-2016].

[39] Bernardo Cuenca Grau and Boris Motik, 'W3C: OWL 2 Web Ontology Language: Model-Theoretic Semantics', *OWL 2*, 2008. [Online]. Available: https://www.w3.org/TR/2008/WD-owl2-semantics-20080411/.

[40] N. M. Franz, R. K. Peet, and A. S. Weakley, 'On the Use of Taxonomic Concepts in Support of Biodiversity Research and Taxonomy', *North*, vol. 76, no. 787, pp. 1–44, 2006.

[41] Merriam-Webster, 'Merriam Webster Dictionay and Thesuarus'. [Online]. Available: http://www.merriam-webster.com/. [Accessed: 04-Oct-2016].

[42] Dieter Fensel, *Ontologies.* Springer Berlin Heidelberg, 2001.

[43] N. Guarino, *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, vol. 46. IOS press, 1998.

[44] U. Hustadt, R. A. Schmidt, and L. Georgieva, 'A Survey of Decidable First-Order Fragments and Description Logics', *Journal on Relational Methods in Computer Science*, vol. 1, pp. 251–276, 2004.

[45] F. Baader, Ed., *The description logic handbook: theory, implementation, and applications*, 2nd ed. Cambridge ; New York: Cambridge University Press, 2007.

[46] I. Horrocks, P. F. Patel-Schneider, and F. V. Harmelen, 'From SHIQ and RDF to OWL: The making of a web ontology language', *Web semantics: science, services and agents on the World Wide Web*, vol. 1, no. 1, pp. 7–26, 2003.

[47] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, 'Methodologies, tools and languages for building ontologies. Where is their meeting point?', *Data & Knowledge Engineering*, vol. 46, no. 1, pp. 41–64, Jul. 2003.

[48] J. De Bruijn, 'Using ontologies: Enabling Knowledge Sharing and Reuse on the Semantic Web', DERI, Technical Report DERI Technical Report DERI-2003-10-29, 2003.

[49] T. R. Gruber, 'A translation approach to portable ontology specifications', *Knowledge acquisition*, vol. 5, no. 2, pp. 199–220, 1993.

[50] S. C. Kleene, *Mathematical logic*. Courier Corporation, 2002.

[51] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler, 'OWL 2: The next step for OWL', *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 4, pp. 309–322, 2008.

[52] M. Krötzsch, *OWL 2 Profiles: An introduction to lightweight ontology languages*. Springer, 2012.

[53] 'OWL @ Manchester: List of Reasoners', 2016. [Online]. Available: http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/. [Accessed: 04-Oct-2016].

[54] W3C, 'W3C Wiki: Ontology Editors', 2015. [Online]. Available: https://www.w3.org/wiki/Ontology_editors. [Accessed: 01-Oct-2016].

[55] 'Open Semantic Framework: Ontology Tools', *Open Semantic Framework*, 2014. [Online]. Available: http://wiki.opensemanticframework.org/index.php/Ontology_Tools. [Accessed: 01-Oct-2015].

[56] M. Krötzsch, F. Simancik, and I. Horrocks, 'A description logic primer', *arXiv preprint arXiv:1201.4089*, 2012.

[57] 'The Protégé Ontology Editor'. [Online]. Available: http://protege.stanford.edu.

[58] R. Shearer, B. Motik, and I. Horrocks, 'HermiT: A Highly-Efficient OWL Reasoner.', in *OWLED*, 2008, vol. 432, p. 91.

[59] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, 'Pellet: A practical owl-dl reasoner', *Web Semantics: science, services and agents on the World Wide Web*, vol. 5, no. 2, pp. 51–53, 2007.

[60] D. Tsarkov and I. Horrocks, 'FaCT++ description logic reasoner: System description', in *Automated reasoning*, vol. 4130, Springer Berlin Heidelberg, 2006, pp. 292–297.

[61] 'Fact++'. [Online]. Available: http://owl.man.ac.uk/factplusplus/.

[62] R. J. Brachman, 'What IS-A is and isn't: An analysis of taxonomic links in semantic networks', *Computer;(United States)*, vol. 16, no. 10, 1983.

[63] J. B. L. Bard and S. Y. Rhee, 'Ontologies in biology: design, applications and future challenges', *Nature Reviews Genetics*, vol. 5, no. 3, pp. 213–222, 2004.

[64] GOC, 'The Gene Ontology', *The Gene Ontology Consortium*, 2015. [Online]. Available: http://www.geneontology.org/.

[65] N. Dessì and B. Pes, 'The Effectiveness of Gene Ontology in Assessing Functionally Coherent Groups of Genes: A Case Study', in *Trends in Applied Knowledge-Based Systems and Data Science*, vol. 9799, H. Fujita, M. Ali, A. Selamat, J. Sasaki, and M. Kurematsu, Eds. Cham: Springer International Publishing, 2016, pp. 283–293.

[66] D. Botstein *et al.*, 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, vol. 25, no. 1, pp. 25–9, 2000.

[67] L. du Plessis, N. Škunca, and C. Dessimoz, 'The what, where, how and why of gene ontology? A primer for bioinformaticians', *Briefings in Bioinformatics*, pp. 723–735, 2011.

[68] S. Y. Rhee, V. Wood, K. Dolinski, and S. Draghici, 'Use and misuse of the gene ontology annotations', *Nature Reviews Genetics*, vol. 9, pp. 509–515, Jul. 2008.

[69] 'SNOMED'. [Online]. Available: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.

[70] Stefan Schulz, 'BioTop Ontology', *BioPortal*. [Online]. Available: https://bioportal.bioontology.org/ontologies/BT. [Accessed: 02-Nov-2016].

[71] E. Beisswanger, S. Schulz, H. Stenzhorn, and U. Hahn, 'BIOTOP: an upper domain ontology for the life sciences', *Applied Ontology*, vol. 3, no. 4, pp. 205–212, 2008.

[72] G. C. Griffiths, 'On the foundations of biological systematics', *Acta biotheoretica*, vol. 23, no. 3–4, pp. 85–131, 1974.

[73] 'Environment Ontology', *BioPortal*. [Online]. Available:

https://bioportal.bioontology.org/ontologies/ENVO. [Accessed: 02-Nov-2016].

[74] B. Bennett, 'Foundations for an Ontology of Environment and Habitat.', in *FOIS*, 2010, pp. 31–44.

[75] NCBI, 'NCBI. National Center for Biotechnology Information: Organismal Classification.', 2016. [Online]. Available: https://bioportal.bioontology.org/ontologies/NCBITAXON.

[76] *Vertebrate Taxonomy Ontology*. .

[77] P. E. Midford *et al.*, 'The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes.', *J. Biomedical Semantics*, vol. 4, p. 34, 2013.

[78] 'OBO Foundry'. [Online]. Available: http://www.obofoundry.org.

[79] NCBO, 'BioPortal, the world's most comprehensive repository of biomedical ontologies.' [Online]. Available: https://bioportal.bioontology.org/. [Accessed: 02-Nov-2016].

[80] W. M. Dahdul *et al.*, 'Evolutionary Characters, Phenotypes and Ontologies: Curating Data from the Systematic Biology Literature', *PLoS ONE*, vol. 5, no. 5, p. e10708, May 2010.

[81] T. A. Dececchi, J. P. Balhoff, H. Lapp, and P. M. Mabee, 'Toward Synthesizing Our Knowledge of Morphology: Using Ontologies and Machine Reasoning to Extract Presence/Absence Evolutionary Phenotypes across Studies', *Systematic Biology*, vol. 64, no. 6, pp. 936–952, Nov. 2015.

[82] NESCent, 'Phenoscape Knowledgebase', *Phenoscape Knowledgebase*, 2017. [Online]. Available: http://kb.phenoscape.org/#/home. [Accessed: 11-May-2017].

[83] C. E. Blank, H. Cui, L. R. Moore, and R. L. Walls, 'MicrO: an ontology of phenotypic and metabolic characters, assays, and culture media found in prokaryotic taxonomic descriptions', *Journal of Biomedical Semantics*, vol. 7, no. 1, Dec. 2016.

[84] C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, and M. A. Haendel, 'Uberon, an integrative multi-species anatomy ontology', *Genome biology*, vol. 13, no. 1, p. R5, 2012.

[85] C. Mungall and M. Haendel, 'Uberon', *Uberon*, 2017. [Online]. Available: http://uberon.org. [Accessed: 05-May-2017].

[86] W. M. Dahdul *et al.*, 'A Unified Anatomy Ontology of the Vertebrate Skeletal System', *PLoS ONE*, vol. 7, no. 12, 2012.

[87] C. Weiland, R. Hoehndorf, G. Gosline, Q. Groom, T. Hamann, and M. Schmidt, 'FLOPO-An Ontology for the Integration of Trait Data from Digitized Floras and Plant Image Collections'.

[88] R. Hoehndorf *et al.*, 'The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants', *Journal of Biomedical Semantics*, vol. 7, no. 1, Dec. 2016.

[89] HPO, 'Human Phenotype Ontology', 2017. [Online]. Available: http://human-phenotype-ontology.github.io/about.html.

[90] S. Köhler *et al.*, 'The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data', *Nucleic Acids Research*, vol. 42, no. D1, pp. D966–D974, Jan. 2014.

[91] S. Köhler *et al.*, 'The Human Phenotype Ontology in 2017', *Nucleic Acids Research*, vol. 45, no. D1, pp. D865–D876, Jan. 2017.

[92] P. N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos, 'The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease', *The American Journal of Human Genetics*, vol. 83, no. 5, pp. 610–615, Nov. 2008.

[93] P. Robinson and S. Mundlos, 'The Human Phenotype Ontology', *Clinical Genetics*, vol. 77, no. 6, pp. 525–534, Jun. 2010.

[94] M. J. Yoder, I. Miko, K. C. Seltmann, M. A. Bertone, and A. R. Deans, 'A gross anatomy ontology for Hymenoptera', *PLoS one*, vol. 5, no. 12, p. e15991, 2010.

[95] HAO, 'Hymenoptera Anatomy Ontology Portal'. [Online]. Available: http://portal.hymao.org/projects/32/public/ontology/.

[96] P. Mullins, R. Kawada, J. Balhoff, and A. Deans, 'A revision of Evaniscus (Hymenoptera, Evaniidae) using ontology-based semantic phenotype annotation', *ZooKeys*, vol. 223, pp. 1–38, Sep. 2012.

[97] K. Seltmann *et al.*, 'A hymenopterists' guide to the Hymenoptera Anatomy Ontology: utility, clarification, and future directions', *Journal of Hymenoptera Research*, vol. 27, pp. 67–88, May 2012.

[98] K. C. Seltmann, Z. Pénzes, M. J. Yoder, M. A. Bertone, and A. R. Deans, 'Utilizing Descriptive Statements from the Biodiversity Heritage Library to Expand the Hymenoptera Anatomy Ontology', *PLoS ONE*, vol. 8, no. 2, p. e55674, Feb. 2013.

[99] Martin Ramirez, 'The Spider Ontology', *BioPortal*, 2017. [Online]. Available: https://bioportal.bioontology.org/ontologies/SPD. [Accessed: 11-May-2017].

[100] W. Wheeler, 'The Tree of Life: Phylogeny of Spiders', *ATOL: Phylogeny of Spiders*, 2017. [Online]. Available: http://research.amnh.org/atol/files/. [Accessed: 11-May-2017].

[101] J. Dönitz *et al.*, 'TrOn: An Anatomical Ontology for the Beetle Tribolium castaneum', *PLoS ONE*, vol. 8, no. 7, p. e70695, Jul. 2013.

[102] C. D. Eardley, 'Afrotropical bees now: what next?', *Pollinating bees: The Conservation Link between Agriculture and Nature. Ministry of the Environment, Brazil. Brasilia, DF*, pp. 105–114, 2002.

[103] Aurona Gerber, Connal Eardley, and Nishal Morar, 'An Ontology-based Key for Afrotropical Bees', in *Proceedings of FOIS 2014, The 8th International Conference on Formal Ontology in Information Systems*, Rio de Janeiro, Brazil, 2014.

[104] M. Horridge, 'Existential Query Plugin', 2017. [Online]. Available: https://github.com/protegeproject/existentialquery. [Accessed: 11-May-2017].

[105] O. Lassila, R. R. Swick, R. Ralph, and others, 'Resource Description Framework (RDF) model and syntax specification', 1998.