



## Ontology-based support for taxonomic functions



Aurona Gerber<sup>a,b,\*</sup>, Nishal Morar<sup>b,d</sup>, Thomas Meyer<sup>b,c</sup>, Connal Eardley<sup>d,e</sup>

<sup>a</sup> Department of Informatics, University of Pretoria, Pretoria, South Africa

<sup>b</sup> Center for Artificial Intelligence Research (CAIR), Cape Town, South Africa

<sup>c</sup> Department of Computer Science, University of Cape Town, South Africa

<sup>d</sup> Department of Computer Science, University of Kwazulu-Natal, South Africa

<sup>e</sup> Agricultural Research Council, South Africa

### ARTICLE INFO

#### Keywords:

Taxonomy  
Systematics  
Taxonomic functions  
Systematics  
Computational ontologies  
Reasoning  
Classification  
Afrotropical bees

### ABSTRACT

This paper reports on an investigation into the use of ontology technologies to support taxonomic functions. Support for taxonomy is imperative given several recent discussions and publications that voiced concern over the taxonomic impediment within the broader context of the life sciences. *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.

### 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 (CHM, 2016; Hester and Harrison, 2007; Stork, 1993). 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 (UNESCO, n.d.; Mora et al., 2011). 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 (Mora et al., 2011).

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 (Guerra-García et al., 2008). 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 (Eardley and Urban, 2010).

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 (Hoagland, 1996). 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 (Hoagland, 1996). This crisis has elevated discussions about

\* Corresponding author at: Department of Informatics, University of Pretoria, Pretoria, South Africa.  
E-mail address: [aurona.gerber@up.ac.za](mailto:aurona.gerber@up.ac.za) (A. Gerber).

computerized support for taxonomy and the lack thereof (Guerra-García et al., 2008; Stork, 1993; Walter and Winterton, 2007; Wheeler et al., 2004). Most major museums house natural collections with actual data and observation records spanning decades, which provide invaluable information for biodiversity conservation (Viscardi, 2016). These collections often also contain unstudied specimens and new species not yet described (Hoagland, 1996). 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 (Godfray, 2002; Godfray et al., 2007). 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 (Maxted, 1992). 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 (Hoagland, 1996). Even though this convention was more than two decades ago,

the taxonomic impediment is just as relevant today, if not more, given the rising concerns for biodiversity protection and sustainable development (Dar et al., 2012; de Carvalho et al., 2007). 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 (Guerra-García et al., 2008; Nickerson et al., 2013; Wheeler et al., 2004). 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 (MacLeod, 2007).

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 initiative with semantic standardization (Bengtson, 1988; Sigovini et al., 2016).

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 (TDWG, 2017). 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 (Deans et al., 2015; Robinson and Webber, 2014). Ontologies, specifically, are often used for such initiatives since an ontology is considered as a *formal, explicit specification of a shared conceptualisation* (Studer et al., 1998). 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 (Walter and Winterton, 2007). 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 (Saupe, 2001). 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 (Guerra-García et al., 2008; Walter and Winterton, 2007). 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). Fig. 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* (Eardley et al., 2010).

Download English Version:

<https://daneshyari.com/en/article/5741908>

Download Persian Version:

<https://daneshyari.com/article/5741908>

[Daneshyari.com](https://daneshyari.com)