Semantic Vernacular System: an Observation-based, Community-powered, and Semantics-enabled Naming System for Organisms

Han Wang[1], Nathan Wilson[2], Kathryn M. Dunn[3], and Deborah L. McGuinness[4]

[1] Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY 12180, USA wangh17@rpi.edu
[2] Marine Biological Laboratory, Woods Hole, MA 02556, USA nwilson@eol.org
[3] Rensselaer Libraries, Rensselaer Polytechnic Institute, Troy, NY 12180, USA dunnk2@rpi.edu
[4] Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY 12180, USA dlm@cs.rpi.edu

Abstract. The Semantic Vernacular System is a novel naming system for creating named, machine-interpretable descriptions for groups of organisms. Unlike the traditional scientific naming system, which is based on evolutionary relationships, it emphasizes the observable features of organisms. By independently naming the descriptions composed of sets of observational features, as well as maintaining connections to scientific names, it preserves the observational data used to identify organisms. The system is designed to support a peer-review mechanism for creating new names, and uses a controlled vocabulary encoded in the Web Ontology Language to represent the observational features. A prototype system is under development in collaboration with the Mushroom Observer website. It allows users to propose new names and descriptions, provide feedback on those proposals, and ultimately have them formally approved. This effort aims at offering the mycology community a knowledge base of fungal observational features and a tool for identifying fungal observations and further provides a prototype example of how other communities may use a semantically-enabled portal to evolve community naming systems.

Keywords: observation, scientific name, fungus, ontology

1 Introduction

With focus on the evolutionary relatedness between and within named groups such as species, genera, families, etc., the precise application of scientific names often requires careful microscopic work, or increasingly, genetic sequencing[10]. This is problematic to many audiences, especially field biologists, who often do not have access to the instruments and tools required to make identifications on a microscopic or genetic basis. Also, while a number of existing web-based biodiversity observation systems (e.g. eBird[9], iNaturalist[3], ArtPortalen[1], Mushroom Observer[4], etc.) have collected tens of millions of occurrence records and linked observations to scientific literature through scientific names, it would be desirable to connect scientific names to precise, accurate and ideally machine-interpretable descriptions compatible with the Semantic Web. Currently there are few widely accepted standards for how to create and curate such descriptions.

In order to address these issues, we proposed the Semantic Vernacular System, a new naming system that is based on observational characteristics, authored in a peer-review context, and encoded in an ontology. We used fungi as a test case for the proposed naming system and developed a prototype on the Mushroom Observer website.

2 An Observation-based Naming System

Modern Scientific names are critical for understanding the biological literature and provide a valuable way to understand evolutionary relationships. To validly publish a name, a description is required to separate the group of organisms this name applies to from those other names apply to. The frequent revision of descriptions due to new evolutionary evidences has lead to a fact that a given scientific name may over time have multiple descriptions associated with it and a given published description may apply to multiple scientific names. Because of this many-to-many relationship between scientific names and descriptions, the usage of scientific names as a proxy for descriptions is inevitably ambiguous.

To resolve this ambiguity, the proposed system creates an additional layer of abstraction by naming the descriptions themselves independently from the scientific names with which they are associated. We refer to these descriptions as Semantic Vernacular Descriptions (SVDs). Each SVD is defined by a set of observational features, and is connected to a number of scientific names which match that definitional feature description. The names of such SVDs will be distinct from scientific names and more closely related to common names, e.g. *PineSpike*, *SmoothCordedPuffball*, *OchreAndRedSpongeTruffle*, etc. SVDs provide shortcuts for expressing a recognizable set of observational features, can be used to drive computer assisted identification, and facilitate the understanding of scientific names based on the observational features of associated SVDs.

3 Community-powered Authoring Mechanism

Similar to scientific names, the proposed system provides mechanisms for encouraging peer-review, preventing the reuse of names, applying any agreed upon nomenclatural rules, and avoiding the unintended re-publishing of existing descriptions. There are two stages in the naming process: "under review" and "approved". Any users can create a new SVD as long as they provide a unique name and a machine-interpretable description using a supported ontology. While the SVD is "under review", other users are free to propose alternative names or descriptions. Once a consensus is reached with sufficient support, the SVD is formally "approved" and cannot be changed by the users. Exactly what is considered sufficient support as well as a formal appeal process for revising an approved definition are expected to be refined as the system is used. The metadata of each proposal, including the proposer information, will be recorded as provenance by the system for evaluating the trustworthiness of the proposal.

Figure 1 presents the prototype interface for proposing an SVD definition. The format for observation input is a set of fungal features together with their values. The system then shows the SVDs matching the input feature description. Users will therefore know whether there are any existing SVDs for their observations. By further clicking on one of the displayed matched SVDs, users can also discover which scientific name(s) potentially apply to their observations.

4 Semantics-enabled Encoding Methods

One of the key components of the proposed system is a controlled vocabulary that describes various observational features including ranges for allowable values. Using the Web Ontology Language (OWL)[5], which has already been actively used by significant parts of relevant biology communities[6], we proposed an expandable ontology for fungal SVDs. Figure 2 illustrates the logic model

of the current ontology. The core entity of the ontology is the `SemanticVernacluarDescription` class which has two attributes `VernacularLabel` and `VernacularFeatureDescription`. The latter encompasses an OWL property restriction that describes the observational features. This part can be substituted for vocabularies representing features of other organisms in different applications.

Being encoded in OWL DL, the vocabulary is machine-interpretable and can be processed by existing computationally efficient algorithms. Using SPARQL[8] queries, RDFS reasoning, and OWL classification inferences, the system not only can return which features are included in each fungal SVD, but is capable of returning SVDs that match observations described using this ontology in essence providing identifications. The system also accepts SVD registrations from users and is able to integrate the new entries into the ontology on the fly. This is enabled by the SPARQL 1.1 Update Service[7], which takes the RDF generated from the user interface and then updates the triple store.

5 Conclusion

The Semantic Vernacular System is an observation-based, community-powered, and semantics-enabled naming system for groups of organisms. This system fills the gap between modern scientific names and observation-based descriptions. It also provides a platform for creating and maintaining machine-interpretable descriptions and observational data. We developed a prototype system on the Mushroom Observer website that builds on a peer-reviewed fungal SVD ontology and empowers computer-aided identification of fungi based on their observational features.

6 Future Work

We plan to allow the system to discover and suggest new SVDs based on patterns in user-provided observational data. We are also working on the system's identification capability so that it can return exact, partial, and near matches. Finally, we will continue with the development of the fungal SVD ontology in collaboration with the community including making connections to various existing biology related vocabularies (e.g. EnvO[2]) to capture and standardize the terminology needed to describe fungal observational features. We expect the evolution of the fungal feature ontology to be community driven. However, the revision process for the ontology will be peer-reviewed and versioned. Thus users will be able to suggest new features or new values for existing features, and possibly use those for describing particular observations. However, they will not be able to use them in defining an SVD on the globally deployed website until they have been formally adopted by the community. We are exploring local definition capabilities for temporary usage during the community review process.

References

1. Artportalen, `http://artportalen.se`
2. The Environment Ontology, `http://environmentontology.org`
3. iNaturalist, `http://www.inaturalist.org`
4. Mushroom Observer, `http://mushroomobserver.org`
5. Owl 2 Web Ontology Language Quick Reference Guide, `http://www.w3.org/TR/2009/REC-owl2-quick-reference-20091027`
6. The Phenotype Ontology Research Coordination Network, `http://www.phenotypercn.org`

7. Sparql 1.1 Update, http://www.w3.org/TR/sparql11-update
8. Sparql Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query
9. Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S.: eBird: a Citizen-based Bird Observation Network in the Biological Sciences. Biological Conservation (142), 2282–2292 (2009)
10. Wilson, N., Dunn, K., Wang, H., McGuinness, D.L.: Application of Semantic Technology to Define Names for Fungi. Tech. rep., Tetherless World Constellation at Rensselaer Polytechnic Institute (2012), http://tw.rpi.edu/web/doc/ApplicationofSemanticTechnologytoDefineNamesforFungi

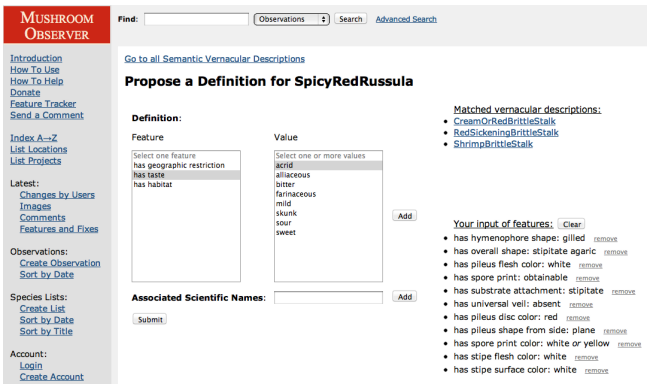

Fig. 1. Interface for proposing a definition for a SVD

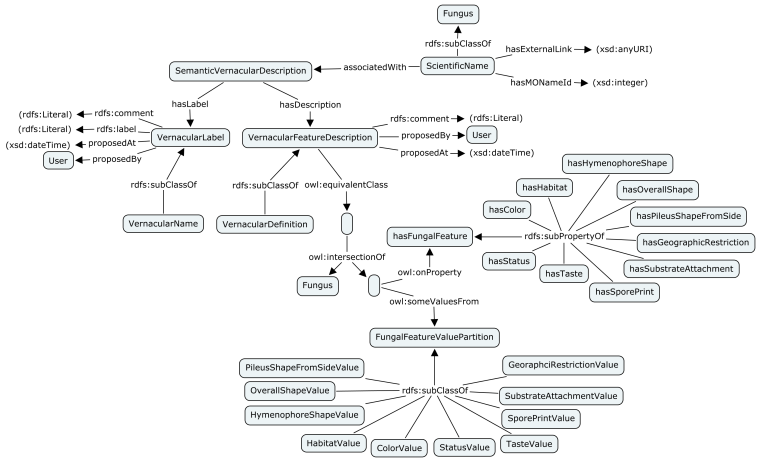

Fig. 2. The logical model of the current fungal SVD ontology