An Ontology of Philippine Natural Products and the API to Retrieve Data from the Ontology

Patrick Ivan S. Altea, Dan Eduard E. Dagdag, Eric Jerome G. Embestro, Nixon Power S. Ong, and Nathalie Rose Lim-Cheng
De La Salle University
** Nathalie Rose Lim-Cheng: nathalie.lim@dlsu.edu.ph*

Abstract: In biochemistry, natural products refer to any chemical compound or substance produced by a living organism. Nowadays, in the pursuit of health (among other reasons), people turn to alternative medicine where the main source are medicinal plants. As a result, there is a bigger demand for scientists to discover inherent chemical compositions of medicinal plants (and other organic matter) that can be used as natural alternatives to chemically synthesized medicines. Apart from identifying the chemical composition of certain plants, they also perform experiments on how to best extract these compounds and test their effectivity in treating certain medical conditions. To record their research findings, they publish results in scientific papers and/or encode these information in databases. Unfortunately, existing online databases lack data about natural products endemic to the Philippines and there are challenges in accessing and updating the entries in the existing works. Some are proprietary, while some do not allow additions or modifications directly. Moreover, those whose data are available are accessible only via searches in web-based tools, rather than having full access to the database. This limits the usability of the data, as software applications cannot be built to maximize the potential of the data. In light of these limitations, an ontology of Philippine natural products has been developed. This paper intends to explain this ontology, as well as the application programming interface (API) developed using the Java language that can be used in the implementation of other software systems to access the data. Validation of the usability of the ontology and the API is done via the development of a simple front-end system to query and update the contents in the ontology.

Key Words: ontology; application programming interface; natural products

1. INTRODUCTION

In the recent decade, health and well-being has become a global trend. With the Philippines having a diverse collection of natural products, the country has seen a rise in the demand as proven by the increase in exportation of these products. The demand locally for organic and natural health products is also increasing. One proof is the doctors' prescriptions of (and the acceptance of the people in taking) malunggay (moringa oleifera) for lactation and lagundi (vitex negundo) for cough, to name a few. As such, scientists and researchers have been looking at natural products, that we use to take for granted, and determining their potential for alleviating certain medical conditions.

There are existing chemical compound databases like PubChem, Super Natural II, and YaTCM, to name a few. Each entry in the database represents a knowledge in the field of chemistry, exhibiting relationship of properties to the natural product. Unfortunately, these databases do not have many of the natural products available in the Philippines. Moreover, directly adding or updating these databases may not be possible. In most cases, the data in these knowledge bases are only accessible via specific search features. This makes exploration of the data for pharmacognosy (i.e., study of medicinal drugs from natural sources) and new drug development difficult.

Therefore, the motivation of our research is to create a database that also contains the natural products found in the Philippines. We designed our database as an ontology, which is a set of concepts on a particular subject area that shows their properties and relations. Maynard et al. (2008) also stated that ontologies are beneficial for health-care and bio-medicine due to the fact that such concepts evolve and update rapidly. Also, as we intend for the data to be used in other software applications, we also designed a set of reusable methods to be part of the API.

First, in section 2, we discuss some of the existing natural products databases. Then, in section 3 we present our ontology of Philippine natural products. Section 4 discusses the API to access the ontology. Validation of the ontology and the API is via the development of a front-end system explained in section 5. Lastly, we conclude in section 6 with our current status and our next targets.

2. EXISTING WORKS

As mentioned, three are several existing natural products databases available. Among which are Super Natural II (Banerjee et al., 2014), Yet another Traditional Chinese Medicine Database (YaTCM) (Li et al., 2018), and NuBBEDB (Pilon et al., 2017).

Super Natural II is an open access web-based database of natural compounds. This database contains information of over 300,000 molecules with its corresponding structures in 2D and in 3D along with its toxicity class. This version can also search for similar compounds by allowing template-based search and provides information on the pathways that are associated with synthesis degradation of the natural products and their mechanism of action. However, this application does not have the feature which can help the users search species and the feature which the user can add or modify entries in its database. It also lacks the natural products that are found in the Philippines.

Similarly, YaTCM is also an open access web-based natural products database that contains over 47,000 natural compounds. Information includes prescription, herbs, ingredients, definite or putative protein targets, pathways, and diseases. The application provides corresponding common names and images of plant-based products and specific plant parts used for treating certain medical conditions. However, this natural products database application only contains natural products that are found in China. It also lacks the feature in which it can accept new entries of natural products.

NuBBEDB is a free online Brazilian natural products database. A validated multidisciplinary information, chemical descriptors, species sources, geographic locations, spectroscopic data (NMR) and pharmacological properties is provided by the database. The application's researchers manually input the data to populate the database, as such this is an incomplete list. The application can search for natural products that are beneficial to certain medical conditions, though it does not specify which part of the natural products must be used.

3. OUR PHILIPPINE NATURAL PRODUCTS ONTOLOGY

The natural products ontology designed is extended from the existing ontology of Philippine medicinal plants by Lim-Cheng et al. (2014) to reduce reduplication of efforts in curating similar / intersecting information that had already been collected and validated. Figure 1 displays the natural products ontology augmented from the medicinal plants ontology. Each encircled entry in the ontology is an entity or concept, while the lines (called edges) represent the relationship between two concepts. Listed below are the entities and their descriptions that we covered in this ontology. Those that are not underlined are from the original Philippine Medicinal Plant Ontology.

(1)MedicinalPlant - name of plants.
(2)Species - scientific name of the plant.
(3)SpeciesPlantPart - plant part of a species
(4)Location - location names where plant was found or is available
(5)Genus - classification of the plant within a family.
(6)Family - names of a family of plants, usually in Latin, where a plant belongs to.
(7)PlantPart - general parts of the plant.
(8)Preparation - preparation procedures.
(9)BodyPart - human body part where the plant part preparation is going to be applied to (or applicable for)
(10)Illness - illnesses or medical conditions
(11)Compound - chemical compounds composing the plant.
(12)CompoundClass - compound families or classification.
(13)BiologicalActivity - beneficial or harmful effects of a chemical compound on an organism.
(14)CellLine - involved cell lines of a biological activity.

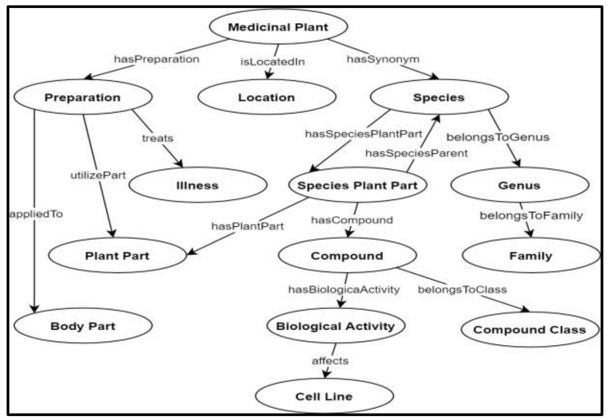

Fig. 1. Natural Products Ontology

Below are the corresponding properties and relationships of the classes defined in the natural products ontology in Figure 1. Those that are not underlined are those from Lim-Cheng (2014).

(1)hasSynonym is the relationship between the MedicinalPlant and Species
(2)isLocatedIn is the relationship of the MedicinalPlant and the location where the plant can be found.
(3)hasPreparation is the relationship between the MedicinalPlant and the steps of preparations to be done in the plant.
(4)belongsToGenus specifies to what Genus the Species belongs to.
(5)hasSpeciesPlantPart is the relationship between the Species and the SpeciesPlantPart
(6)hasSpeciesParent is the relationship between the SpeciesPlantPart and the Species
(7)hasPlantPart is the relationship between the SpeciesPlantPart with the PlantPart that can be used
(8)hasCompound is the relationship indicating what natural compound composes the SpeciesPlantPart.

(9)belongsToFamily specifies to what Family the Genus belongs to.
(10) treats specifies what Illness the Preparation is for
(11)appliedTo specifies what BodyPart is the Preparation applied to.
(12)utilizePart specifies what PlantPart is utilized in the Preparation.
(13)belongsToClass refer to what CompoundClass the Compound belongs to
(14)hasBiologicalActivity specifies the BiologicalActivity of a Compound.
(15)hasCellLine specifies which CellLine the BiologicalActivity affects.

Faculty members from the Chemistry Department and Biology Department of De La Salle University validated the correctness of the ontology design.

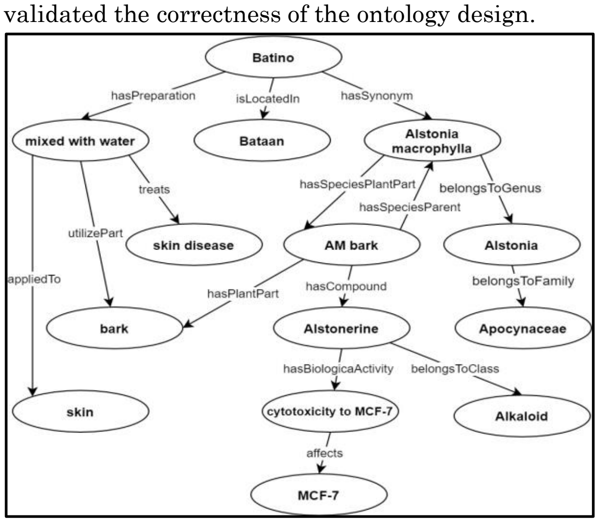

Fig. 2. Sample Natural Products Ontology

To further illustrate the ontology design, we show the entries to be stored in the corresponding concepts the plant Batino (scientific name "Alstonia macrophylla") in Figure 2.

4. THE API TO ACCESS THE ONTOLOGY

The API which will handle the accessing of the ontology is written in Java which utilizes an existing library called Protege which handles the OWL or Web Ontology Language where the ontology of natural products was implemented. The available functionality is intended to be simple and granular to lend reusability to the available features.

4.1 Initialization for Querying

In order to utilize all the features or functions of the program, the class should be first initialized and instantiated by specifying the local path of the ontology to be accessed and creating an object of the OntoQuery class. The created object can now be used to call the features and functions of the class. Figure 3 indicates the constructor signature to initialize/connect to the API.

```
public OntoQuery() throws OntologyLoadException
```

Fig. 3. OntoQuery Constructor

4.2 Retrieving All Medicinal Plant Names

This feature will enable the system for a function call which will obtain all the names of medicinal plants stored in the ontology. As shown in Figure 4, this function returns a list of strings of all the medicinal plant names.

```
public List<String> getAllMedPlantNames()
```

Fig. 4. Get Medicinal Plant Names

4.3 Retrieve Location of a Plant

This feature allows the user to retrieve the location of a specific plant. The method call to the given method in Figure 5 returns a list of strings of all locations by passing a parameter of medicinal plant name of string type.

```
public List<String> getLocations(String MedicinalPlant)
```

Fig. 5. Get Location/s of a Plant

4.4 Retrieve Synonym of a Plant

This feature allows the user to retrieve the synonym or scientific name of a specific plant. As shown in Figure 6, this function returns a list of strings of all synonyms of a plant by passing a parameter of medicinal plant name of string type.

```
public List<String> getSynonyms(String MedicinalPlant)
```

Fig. 6. Get Synonym/s of a Plant

4.5 Retrieve Compounds of a Plant

This feature allows the user to retrieve compound/s found in a specific plant. As shown in Figure 7, this function returns a list of strings of all compounds that are found in a plant by passing a parameter of medicinal plant name of string type.

```
public List<String> getCompounds(String MedicinalPlant)
```

Fig. 7. Get Compound/s in a Plant

4.6 Retrieve Medicinal Plant Object

This feature allows the user to retrieve the object of a medicinal plant including its other related classes like the taxonomic information, location, synonym, different parts of the plant, compounds found in the plant, and all the other classes in the ontology. As shown in Figure 8, the function returns an object composing of its different classes by passing on the plant name of string type.

```
public List<MedicinalPlant> getMedPlant(String MedicinalPlant)
```

Fig. 8. Get Medicinal Plant Object

5. SIMPLE APPLICATION AS VALIDATION OF API

A simple front-end system was developed as proof that the methods available can be used to retrieve data, as well as, update the contents of the ontology. In this application, the following features are available for the user:

1. Add - The user can manually add a plant entry in the system.
2. Modify - The user can modify specific information in the plant entry in the database.
3. Search - The user can search either plants, genus, family, chemical compounds and other concepts for faster viewing of entries which the user may need. A sample search result page as shown in Figure 9 returns a list of all medicinal plant names in the ontology matching the keyword entered by the user within the corresponding selected category.

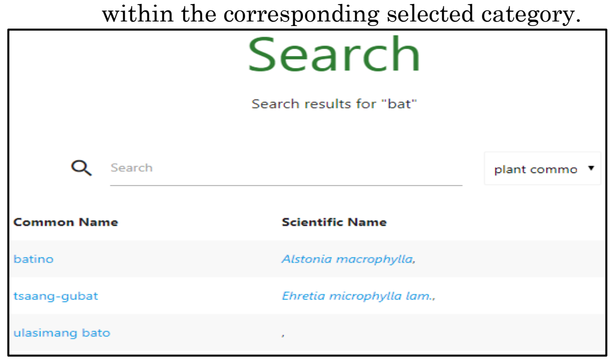

Figure 9: Search Result Page

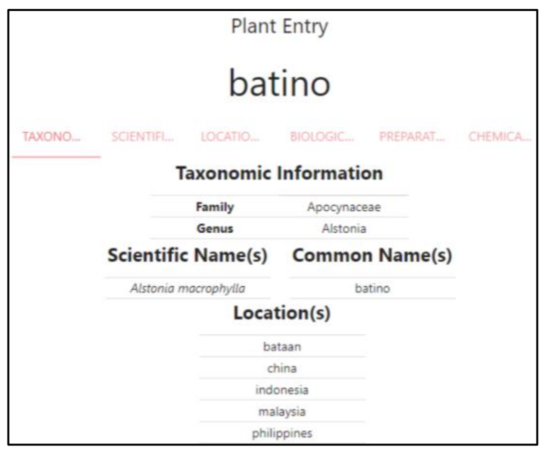

Figure 10: View Entry Page

On the other hand, calling the function getMedPlant() explained in Section 4.6. will produce the result shown in Figure 10.

4. View - The user can then view entries from the database, whether it is from the search results or from manually viewing the entries themselves as shown in Figure 10. Viewing of an entry will allow the user to view other

related information like the taxonomic information, scientific/common names, locations, biological activities, preparations, and chemical compounds found in the plant.

Based on functional testing, all the methods produce the correct results and the API proves sufficient in accessing all available information. We have also had the features reviewed by target end users and these have been received well. Though they are requesting for additional advanced search functionalities to be included in the final system.

6. CONCLUSION

Existing ontologies and databases cannot be populated with new entries directly. Either they are proprietary, including the data, or they need data to be submitted for review prior to their addition. This prevents users of these information to directly modify and access these information. Moreover, as there is no direct way to access these, they cannot be used directly in software systems. We presented in Section 3 our ontology design (those underlined which are) the new concepts and relationships that were incorporated from the work of Lim-Cheng, et al (2014). The methods in the API used in creating the application discussed in Section 5 can be used as part of the implementation of other applications like recommendation systems that would suggest recipes that include natural products that contain chemical compounds known to alleviate the user's medical condition, among others.

Currently, we are augmenting the application to address the request for advanced search features. These can be implemented from the existing API already developed. More extensive user testing will be performed consequently.

The application presented (and the API) is part of a research that will automatically populate the ontology from scientific publications (rather than existing databases) using natural language processing techniques. This would alleviate the need for a fully manual encoding of the data, as well as making sure that new findings in reputable publications are considered in the data stored.

7. ACKNOWLEDGMENTS

We would like to express our gratitude to faculty members from the College of Science, namely Dr. Angelyn R. Lao, Dr. Raymond S. Malabed, Mr. Vincent Antonio S. Ng, and Dr. Maria Carmen Tan for their inputs and feedback on the ontology design.

8. REFERENCES

Banerjee, P., Erehman, J., Gohlke, B., Wilhelm, T., Preissner, R., & Dunkel, M. (2014, 10). Super natural ii–a database of natural products. Nucleic acids research. doi: 10.1093/nar/gku886

Li, B., Ma, C., Zhao, X., Hu, Z., Du, T., Xu, X., Lin, J. (2018, 11). Yatcm: Yet another traditional chinese medicine database for drug discovery. Computational and Structural Biotechnology Journal, 16. doi: 10.1016/j.csbj.2018.11.002

Lim-Cheng, N. R., Co, J. R., Gaudiel, C. H. S., Umadac, D. F., & Victor, N. L. (2014). Semi-Automatic Population of Ontology of Philippine Medicinal Plants from On-line Text. In DLSU Research Congress, De La Salle University, Manila, Philippines.

Maynard, D., Li, Y., & Peters, W. (2008). NLP techniques for term extraction and ontology population

Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., Bolzani, V. S. (2017). Nubbedb: an updated database to uncover chemical and biological information from brazilian biodiversity. Scientific Reports, 7 (1), 7215. doi: 10.1038/s41598-017-07451-x