A Logical Model for Taxonomic Concepts for Expanding Knowledge using Linked Open Data

Rathachai Chawuthai[1], Hideaki Takeda[2], Vilas Wuwongse[3], and Utsugi Jinbo[4]

[1] Asian Institute of Technology, Prathumtani, Thailand
rathachai.chawuthai@ait.ac.th
[2] National Institute of Informatics, Tokyo, Japan
takeda@nii.ac.jp
[3] Thammasat University, Prathumtani, Thailand
wvilas@engr.tu.ac.th
[4] National Museum of Nature and Science, Tokyo, Japan
ujinbo@kahaku.go.jp

Abstract. The variety of classification systems and the new discovery of taxonomists lead to the diversity of biological information, especially taxon concepts. The association among taxon concepts across research institutes is very difficult to establish, because there is no single interpretation of the name of a taxon concept. Owing to this difficulty, further expansion of more biological knowledge is very complicated for taxonomists when they deal with many sources of data or ambiguous concepts. In order to link some relevant taxon concepts across research repositories, it is necessary to consider the precise context of bioloical data also. As a result, we propose a logical model for taxon concepts in Resource Description Framework (RDF). Moreover, we implement a prototype to demonstrate the feasibility of our approach. It has been found that our model can publish taxon information as linked data and, hence, with additional benefits from Linked Open Data (LOD) cloud.

Keywords. Logical model, Linked data, Ontology, Taxon Concept

1 Background

More than 1.4 million species throughout the world have been truly described and classified with appropriate naming depended upon their characteristics; such as, morphological characters, living behaviors, DNA sequences, etc. [1-2]. Many taxonomists have studied living things, research, and publish their knowledge for over hundred years. However, before the information age, their researches were not completely shared across all researchers around the world. Some researchers might have different perspectives to classify and name their specimens. It is possible that the same species may be classified and named differently [2]. For example, *Papilio xuthus*, Chinese yellow swallowtail butterfly, has been given at least four names by four taxonomists. As a result, scholars and researchers are inconvenient to study all information about one living thing completely from a single taxonomic name.

2 The Proposed Logical Model

Our paper presents a logical model and ontology for linking taxon concepts which comprises a series of changes, the diversity of taxonomic classifications, and the variety of naming. For the purpose of linking data, we have developed our model by employing ontology of contextual knowledge evolution together with some widely accepted ontology such as LODAC [3] and SKOS [4]. This research proposed some useful operations that specify the changes of taxon concepts; for instance, change taxonomic hierarchy, rename, merge, replace, split, etc. We also introduced some kinds of links between taxon concepts, for example, common name, correct spelling, homonym, junior synonym, senior synonym, etc. The model is expressed in an ontology named Linked Taxonomic Knowledge (LTK). Moreover, we enhance the Contextual Knowledge for Archives (CKA) ontology in order to deal with both dynamic and static information represented in RDF and hence the history of the taxon concept can be traced back [5]. For example, two genera of owls named *Nyctea* and *Bubo* have been merged into the latter genus *Bubo*. Following the change of genera, the scientific name of a snowy owl *Nyctea scandiaca* has been replaced by *Bubo scandiacus* in order to satisfy the convention of scientific name [6-7]. These facts will be presented in RDF that satisfies the logical model from the CKA approach as follows:

```
ex:even1999     cka:interval          [tl:beginAtDateTime "1999"] ;
                cka:assure            ex:mg1, ex:rp1 .
ex:mg1          rdf:type              ltk:MergingTaxonConcept ;
                cka:conceptBefore     genus:Bubo, genus:Nyctea ;
                cka:conceptAfter      genus:Bubo_1999 .
ex:rp1          rdf:type              ltk:ReplacingTaxonConcept ;
                cka:conceptBefore     species:Nyctea_scandiaca ;
                cka:conceptAfter      species:Bubo_scandiacus .
```

In practice, we implement prototype by having inference rule, so it can trace back to the changes of these species' names that are caused from the merging of the two genera. In addition, the application provides links to the relevant data with the data in LODAC. As a result, we found that our LTK model is feasible and suitable for collecting the change of taxon concepts and establishing links to relevant concepts across research institutes in order to expand more knowledge about taxon concepts.

References

1. Darwin, C., Peckham, M.: On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. Penn Press, Philadelphia (1959)
2. Winston, J. E.: Describing Species: Practical Taxonomic Procedure for Biologists. Columbia University Press, New York (1999)
3. Linked Open Data for Academia, http://lod.ac/
4. http://www.w3.org/2009/08/skos-reference/skos.html
5. Chawuthai, R., Wuwongse, V., Takeda, H.: A Formal Approach to the Modeling of Digital Archives. In: ICADL'12, Taipei (2012)
6. Wink, M., Heidrich, P.: A Guide to Owls of the World. Yale University Press (1999)
7. International Code of Zoological Nomenclature, http://www.iczn.org