Temporal Versioning in Biodiversity Ontology

Andini Sintawati, Lintang Yuniar Banowosari, I Made Wiryana, I Wayan Simri Wicaksana,
Universitas Gunadarma Jakarta, Indonesia
@staff.gunadarma.ac.id, @staff.gunadarma.ac.id, mwiryana@staff.gunadarma.ac.id, @staff.gunadarma.ac.id

***Abstract* — The Indonesian archipelago has an extraordinary biodiversity, and it has dynamic and potentially informative biogeography. Biodiversity Information requires the combined data on living creatures and their habitats, build a model that connects all kinds of information. Heterogeneous data handled, provided and distributed by different research groups, which collects data using a different vocabulary. Ontology adopted as one way to reduce the problem of heterogeneity, thus helping the cooperation between researchers. Thus, in order to provide co-operation group, several types of ontology integration mechanism must be provided. An interface (mediator) to various sources of data needed to integrate information from sources that are used, either by human users or applications programs. Interface (in the form of applications) determine the source of data used and how to get back data from the data source.**

Keywords: Biodiversity, ontology, interoperability, integration ,temporal versioning

I. INTRODUCTION

The Indonesian archipelago has a biodiversity that extraordinary, and arguably the region's most dynamic and potentially informative biogeography. If a species with the same features found in other areas and in fact these two species have a relationship then that information will be viewed through the mapping if the species are in the same hierarchy.

Information of biological diversity requires a combination of data on living creatures and their habitats, build a model that connects all kinds of information. Heterogeneous data handled, provided and distributed by different research groups, which collects data using a different vocabulary. Ontology adopted as one way to reduce the problem of heterogeneity, thus helping the cooperation between researchers.

Exchange of information and integration of information between systems is known as interoperability. Interoperability becomes more important when the system of interacting multiply. Semantic data integration is the process of using a conceptual presentation of data and their relationships to eliminate the possible heterogeneity. Heart semantic data integration is the concept of ontology, namely a specification of explicit of a shared conceptualization [3]

II. ONTOLOGY

One of the most widely used definition of ontology is [3]:"ontology is an explicit specification of conceptualization". From the perspective of computer science, an ontology can be viewed as a data model that represents a set of concepts in the domain and the relationships between the concepts. Knowledge in ontologies is formalized using four types of components:

- Classes: sets, or types of objects (concepts or categories of concepts in the domain), usually organized in taxonomies;
- Instances: objects in the domain, represented as an instance of class;
- Properties: used to describe examples / class. Properties to express the attributes of objects or relations;
- Constraints: an abstraction that uses properties to describe the class.

In distributed and open systems, ontology alone can not solve the problem of interoperability and heterogeneity. Different research groups may have different interests, the purpose of research, use tools and manipulate a variety of computing knowledge at various levels of detail and abstraction. Thus, in order to provide co-operation group, several types of ontology integration mechanism must be provided.

Approaches to start the integration of two ontologies ontology o1, and o2:

- Mapping: preprocessing stage, to identify all the concepts in the o1 and o2 are identical, using appropriate techniques;
- Merge: build a new ontology based on the mapping between o1 and o2, merge equivalent concepts into new concepts. This concept is receiving a name derived from the concept of o1 or o2;
- Alignment: build a new ontology that match and maintain the original ontology, which are related according to the mapping detected.

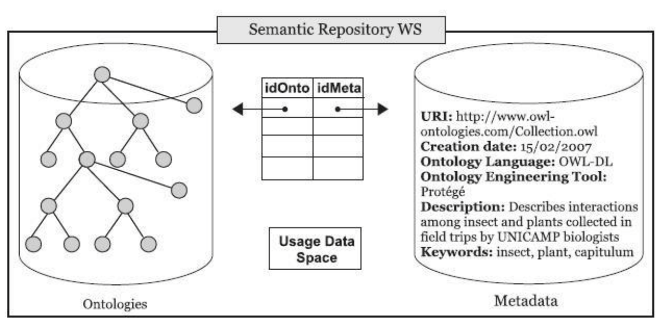

Figure 1. Examples of Semantic Repository

Figure 1. describes the Semantic Repository structure, prepared by the ontology and metadata data space, and space usage data. This figure shows that every ontology is stored has an associated metadata structure, showing, for example, URI, date of creation, and related keywords.

III. METHODOLOGY

Biodiversity research is a multidisciplinary field that requires the cooperation of various scientists who collect, correlate and analyze data on living creatures and their habitats. They collect data using a different vocabulary, in many formats and participated in various standards. Ontology adopted as one way to reduce the problem of heterogeneity, thus helping the cooperation among researchers. Hence, interoperability and heterogeneous data manipulation is one of the major challenges faced by scientists.

An interface (mediator) to various sources of data needed by the human or applications programs. Interface (in the form of applications) determine the source of data used and how to get back data from the data source. This interface integrates information from sources that are used by the user.

The integration of information is the problem of combination of data residing in different sources, and provides users with a schema that holds this data. This chart is commonly called Common ontology, is a customized view of information.

Exchange of information and integration of information between systems is known as interoperability. Interoperability becomes more important when the system of interacting multiply. Semantic data integration is the process of using a conceptual presentation of data and their relationships to eliminate the possible heterogeneity. Heart semantic data integration is the concept of ontology, namely a specification of explicit of a shared conceptualization.

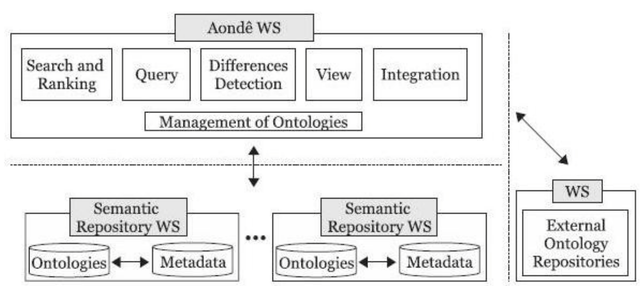

Figure 2. Two-tier architecture to manage the ontology, separate the persistence layer of semantic operations Aonde

Figure 2 Shows the basic architecture, where ontology stored in several distributed ontology repositories, each accessed through a web service. This repository can be of two types: Repository Semantic built and managed by the Semantic Repository service, and external third parties who publish the data repository Ontology through WSS. Aonde provides an extensible set of modules that can be called by client applications to search, rank, query, integrate, create views and comparing ontology.

World Wide Web (WWW) is the greatest means of information exchange and open to all, anyone with an Internet connection. Advantages WWW has attracted much interest agencies to distribute their information on the internet, including biodiversity. The need for the distribution of biodiversity information through the internet media has been very necessary. Distribution of information through the Internet using the World Wide Web are expected to disseminate information with greater reach and unlimited time.

Currently people have to manually visit multiple websites to find biodiversity data they want. They must collect the data from the same species from several websites, spend a long time to collect such data.

The focus of this research is to provide concepts, mechanisms of mapping for heterogeneous data sources for Facebook to integrate the semantic level, which in turn allows the existing data on the sources of heterogeneous data so that semantic data integration can be achieved.

Finding a solution to solve the problem of semantic diversity among sources of heterogeneous data, the mapping required to convert the terminology known in their respective data sources and create algorithms for mapping process for mapping can be performed according to the existing terminology in their respective data sources.

IV. APPROACH

Ontology mapping areas of research related to both taxonomy and merging dictionaries / encyclopedias. Doerr (2001) explains the semantic difference between the dictionary / thesaurus that affect the mapping process. Widerhoed (1994) describes the composition of the ontology and domain ontology difference. By generalizing and adding other cyc typical semantic problem, obtain the following list of differences between source ontology and the ontology of Cyc (in order of complexity):

- Differences terminology
  - ✓ Different names for the same concept
  - ✓ Related but different concepts
  - ✓ More specialized or general version for the same concept
  - ✓ Attributes vs. functions vs. predicates representation
- Differences simple structure
  - ✓ Two ontologies are similar, but not yet fused
  - ✓ One ontology is a subset of the other
  - ✓ One ontology is a reorganization of the other
- Differences complex structure
  - ✓ For instance having action predicates vs. reified events
- Representation basic difference
  - ✓ For example, Bayesian probability logic vs. truth

Mapping the term (for the Cyc ontology) form the majority of ontology mapping experience. Source ontology typically contains a finite set of terms that have well-defined attributes. The source term used in mapping taxonomies usually but not always the case.

Federal Information Processing Standard (The Federal Information Processing Standards (FIPS)) is a set of standards that describe document processing code. They map the FIPS codes that designate non-US countries and their main administrative division into Cyc, creating a new term for the geographic entity that does not exist in Cyc. This is the simplest form of mapping ontologin which mapped a single relationship, and where the word is missing in the reference ontology Cyc can be easily identified and made.

WordNet has become a standard lexical knowledge base with more than 130,000 English words and phrases arranged in a taxonomy by the parts of speech. The words are grouped into sets of synonyms (synsets) and assigned ID. Like the Open Directory, synset id change when a new version of ontology is released, but the program is provided for backward compatibility utility mapping synsets between versions. We have mapped more than 12,000 WordNet Cyc term for version 1.6 and continue to support the mapping of WordNet through graphical tool built into Cyc. Below is an example of mapping of WordNet noun synset, synset adverbs, and verb synset:

```
(synonymousExternalConcept
       ShoppingMallBuilding WordNet-1997Version
   "N03144979")

(synonymousExternalConcept West-Generally
       WordNet-1997Version "R00318751")

(synonymousExternalConcept
       (TransportViaFn RoadVehicle)
   WordNet-1997Version "V01317106")
```

In the statement above, functional expression (TransportViaFn RoadVehicle) means the collection of events in which the vehicle road transport (more or less conventional) is a transportation device. Cyc term functional notation is often used as a source ontology mapping so that new concepts can be formed by arranging the concepts that already exist, rather than creating a new reified term for each.

The research in ontology versioning and evolution has borrowed many ideas from schema versioning and evolution in database research. In addition, ontology versioning and evolution are still in their early stages. Thus, it might be beneficial to look at the analogous solutions proposed for the database schema versioning and evolution first before tackling the same problem in ontologies.
Database systems are rarely stable following the initial implementation. Modifying the database schema is a common but often troublesome occurrence in database administration. Schema versioning and evolution arose in the context of long-lived database applications, where stored data were considered worth surviving changes in the database schema [9].

Since the introduction of schema change facilities in the database systems, two fundamental problems were solved: **semantics of change** and **change propagation** [9]. **Semantic of changes** involves the checking and maintenance of schema consistency after changes. Most existing approaches that address the semantics of change use *variants* to define the consistency of a schema, and definite *rules* that check the consistency of those *variants* after each change performed on the schema. In addition, a set of *axioms* are used to formalize the dynamics of schema evolution which is the actual management of schema changes in a system in operation. **Change propagation** involves the consistency of extant data with the modified schema.

Research on ontology versioning and evolution focuses on the issues of how ontologies deal with the internal and external changing environments. According to [10], ontology versioning and evolution is defined as *"the ability to manage ontology changes and their effects by creating and maintaining different variants of the ontology"*. This ability includes methods to distinguish and recognize versions, specifications of relationships between versions, update and change procedures for ontologies, and access mechanisms that combine versions of an ontology and the corresponding data.

While there are many ontology tools/systems providing logs of changes between various versions to support ontology versioning and evolution processes, there is no interaction or sharing of information among these tools/systems. Having a general framework for ontology versioning and evolution that allows tools supporting different evolution tasks to share change information and leverage change information obtained by other tools, will make ontology versioning and evolution process much more efficient [11].
PROMP is a plug-in suite for Protégé used to manage multiple ontologies. It has four main functions:

- **Compare** the current ontology to a different version of the same ontology;
- **Move** frames between the current including project and one of the included projects;
- **Merge** two ontologies and added the resulting merged ontology to the current project;

- **Extract** a portion of another ontology and add it to the current project

REFERENCES

[1] Jaudete Daltio_, Claudia Bauzer Medeiros, Aondeˆ : An ontologi Web service for interoperability across biodiversity applications, Information Systems 33 (2008) 724–753,

[2] Eugene Ch’ng, An Efficient Segmentaion Algorithm for Entity Interaction, Biodiversity Informatics, 6, 2009, pp. 5-17.

[3] Laura L. Downey dan Deana Pennington, Bridging The Gap Between Technology and Science with Examples from Ecology and Biodiversity, Biodiversity Informatics, 6, 2009, pp. 18-27.

[4] Miguel A. Fernandez, Stanley D. Blum, Steffen Reichle, Qinghua Guo, Barbara Holzman, and Healy Hamilton, Locality Uncertainty and The Differential performance of Four Common Niche-Based Modeling Techniques, Biodiversity Informatics, 6, 2009, pp. 36-52.

[5] Stephen L. Reed and Douglas B Lenat, Mapping Ontologies into *Cyc*, *Cyc*orp, Inc, 3721 Executive Center Dr, Austin Texas, 2002.

[6] Lily Wulandari, Interoperabilitas Data Menggunakan Query Rewriting Berbasis Ontologi dan Mapping, Disertasi, Universitas Gunadarma, 2009.

[7] I Wayan Simri Wicaksana. Survei dan Evaluasi Metode Pengembangan Ontologi (Survey and Evaluation of Ontologi Development) Methodology. In Proc. of KOMMIT 2004, Jakarta&Depok-Indonesia, 24 August 2004. Gunadarma University, p. ISSN: 1411-6286,.