

ONTOLOGY-BASED DATA WAREHOUSE VISUALIZATION USING PROTÉGÉ FOR CORONA VIRUS SPREAD INCIDENCES

Tazkia S. Ardan¹, Denada F. Zahra², Meinawati³

¹ Faculty of Computer Science Subang University, ² Faculty of Informatics Management AMIK-YPAT
Purwakarta,

³ Master of System Information Study Program, STMIK LIKMI Bandung, Indonesia
email: tazkiaardan@unsub.ac.id¹, denada.zahra@gmail.com², meinawatiwijaya@gmail.com³

ABSTRACT

Corona Virus (Covid-19) is an infectious disease caused by a newly discovered coronavirus. At this time, there are no specific vaccines or treatments for COVID-19. death rates due Covid-19 in Indonesia reach 8,9%. The presence of differences between data released by the Covid-19 complaint website of West Java Province and data released by Subang Government indicates that's still difficult to obtain accurate information. We need to build data warehouse and create ontology of Covid-19 incidences data to figure out the important information in Covid-19. we are using Protégé to build and visualize ontology of Covid-19 data.

Keywords : ontology, Protégé, Covid-19

1. INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it's important that you also practice respiratory etiquette (for example, by coughing into a flexed elbow). At this time, there are no specific vaccines or treatments for COVID-19. However, there are many ongoing clinical trials evaluating potential treatments. According to WHO data, As of 27 April 2020, a total of 2 878 196 people have been reported confirmed for coronavirus disease (COVID-19) globally. Among these, there have been 198 668 deaths reported related to COVID-19.

The number of positive cases of COVID-19 in Indonesia reached 9,096, as many as

12.65% were recovered and 8.41% of the positive cases had been obtained worldwide. According to the data from the state-run Pikobar, in Subang Regency, dated April 28, 2020, it is positive 15, PDP 12, ODP 952. This data does not include the data in the regency Subang. Most of these endemic data are not organized well. Therefore, it was very difficult for the ministry of health to adapt the effective way for reducing and preventing Covid-19 spread.

A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications. A data warehouse centralizes and consolidates large amounts of data from multiple sources. Its analytical capabilities allow organizations to derive valuable business insights from their data to improve decision-making.

Stanford University specializing in Ontology, has developed software called

Protégé. Protégé is an application used to easily process data with a specified format so that it can easily understand and transformed into a more useful form of data. Then, this data can be utilized in supporting decision making process.

In this paper, we create a visualization of Covid-19's cases data warehouse in Subang using various variables, such as class, entity, object, data property, and individuals.

Data warehouse is usually developed to put together data from various sources and then process it for decision making purpose. A big amount of data and complex relationship among entities in data warehouse make it difficult to understand and generate meaningful information which represents those data. Therefore, we need a clear and detail representation that visualize all of these entity relationship. One way to easily represent the connectivity among entities is by creating the ontology visualization of these entities. The structure of data warehouse should consist of multidimensional models, Online Analytical Processing (OLAP) model for analysis task. Inputs for data warehouse are data that coming from operational database or even data from various web pages. The bigger amount of data processed in data warehouse, then the multidimensional representational would not be a good option because of the complex visualization.

The Semantic Web is an extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C). The goal of the Semantic Web is to make Internet data machine-readable. To enable the encoding of semantics with the data, technologies such as Resource Description Framework (RDF) and Web Ontology Language (OWL) are used. These technologies are used to formally represent metadata. For example, ontology can describe concepts, relationships between entities, and categories of things. These embedded semantics offer significant advantages such as reasoning over data and operating with heterogeneous data sources.

Many researches had utilized ontology for visualizing, digitizing process works, and describing detail of variables and data type in

data warehouse. Insight of our research is quite similar with the notion of the all mentioned research works above, which is to present entities relationship in the data warehouse with a better visual that can be easily understood and used by users. Here, we visualize the relationship of entities about Covid-19 Incidence Spread in Subang.

2. METHOD

1. Ontology

Ontologies are used to capture knowledge about some domain interest. An Ontology describes the concept in the domain and also the relationship that hold between those concepts. Ontology learning (ontology extraction, ontology generation, or ontology acquisition) is a subtask of information extraction. Ontology learning can be made from unstructured sources (NLP, techniques), semi-structured sources (XML, HTML) and structured data (SNOMED CT). Guarino and Giaretta describe that Ontology as a philosophical discipline which deals with the nature and organization of reality [1] with goal to semi-automatically extract relevant concept and relation from a given corpus or other kinds of data set to form Ontology.

The artificial-Intelligence literature contains many definition of an ontology; many of these contradict one another. As explained in [2], an ontology is a formal explicit description of a domain, consisting of classes which are the concept found in the domain (also called entities). Each class may have one or more parent classes, formulating thus a specialization/generalization hierarchy, has properties (also called roles or attributes) describing various features of the modelled concept, and restriction on the slots called facets or role description. Allowed classes for slots of type Instance are often called the range of the slot. The classes to which a slot is attached are called the domain of the slot. Classes may have instances, which correspond to individual objects in the domain of discourse; each instance has a concrete value for each property of the class it

belongs to. An ontology together with a set of individual instances of classes constitutes a knowledge base.

Numerous ontologies have been developed and used in various research areas. Each development project usually follows its own set of principles in order to design criteria and phases in the ontology development process. An ontology development methodology comprises a set of established principle, process practice, methods, and activities used to design, construct, evaluate, and deploy ontologies [3]. Meanwhile common components of Ontologies are: individuals, classes, attributes, relation, function terms, restrictions, rules, axioms, and events [4].

An obvious approach is to structure the ontology as a tree, based on the is-a links and visualize this tree, while role relation is made available only through the respective slots. Some interesting ontology management tool surveys may be found in the Protégé [5]. Various ontology extension has been incorporated into ontology management tools, as is the case of relationship class in Protégé, which is effectively a role relationship that is allowed to have properties [5]. The purpose of this work is to visualization data warehouse with different characteristics and investigate their effectiveness.

2. Protégé

Protégé is the latest in a series of tolls developed in our laboratory to assist users in the construction of a large electronic knowledge bases [6]. The Protégé system presents the user with a series of “tabs,” where each tab offers the user access to a different element of the system’s functionality. Standard tabs allow users to edit and browse a domain ontology, to custom-tailor the knowledge-acquisition tool generated from a domain ontology, to enter knowledge into the knowledge acquisition tool, and to search the knowledge base.

An essential goal of the system is to make knowledge browsing and entry as simple for users as possible. When the system generates a knowledge-acquisition tool from an ontology, users enter domain information by filling in the blanks of intuitive forms, selecting items from lists, and by drawing diagrams. Furthermore, perhaps, one of the biggest values of the Protégé system is the availability of dozens of plugins [7] (developed by the Protégé group and by contributors from all over the world). These plugins provide alternative visualization mechanisms, enable management of multiple ontologies, including merging and version management, allow the use of various inference engines and problem solvers with Protégé ontologies, and provide other functionalities.

Protégé provides a Java API for application developers to access and modify all aspects of Protégé knowledge bases and its user interface. Protégé stores ontologies in many different formats including relational databases, UML, XML, and RDF. We are currently working on providing support for the OWL language, which is a language designed for the next generation of the World-Wide Web-the Semantic Web [8].

3. Research Method

Ontology Philosophical understanding of ontology is basically a study of something that exists. In other words, ontology is a theory of the meaning of an object, the property of an object and its relationships that may occur within the scope of knowledge. Basic Concepts of ontologi, that is :

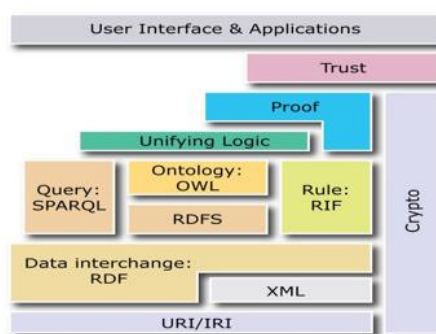


Figure 1. Layers of Ontology

In the paper [9] ontology explains the concept of domains (class, also called concepts). The property of each concept explains the description and attributes of the concept (slots, also called properties), and restrictions on the slot are called facets or boundaries). An ontology together with individual instances of events forms the basis of knowledge. Components in ontology consist of instances, slots and classes.

Class describes the concept of the domain. For example, a class of winners represents all wines. A class can have subclasses that explain more specific concepts from superclass. For example, class wine is divided into whitewine, redwine, and rosewine

Figure 2. Data Covid-19 on MS. Excel

Below are some components that usually employed to form the structure of ontology:

- XML (Extensible Markup Language) provides output syntax for a structured document, but has not been enforced for XML documents using semantic constraints.
- XML Schema is used to confine the structure of XML document.
- RDF (Resource Description Framework) gives a simple semantic of data model of objects or resources and their relationship. It can be expressed using XML syntax.
- RDF Schema is a vocabulary to describe the properties and classes of sources by using a semantic to spread the hierarchies.
- OWL (Ontology Web Language) adds some vocabularies to explain about properties and classes.

OWL As previously described, if you build an ontology, of course, will produce OWL, as a form of ontology that has been built. OWL itself has 3 sub-languages whose usage is different, according to the user's wishes, that is:

- OWL Lite For use based on certain classifications and restrictions, making it easier for users to implement it
- OWL DL Supports maximum use of expressions accompanied by precise calculations, for certain restrictions.

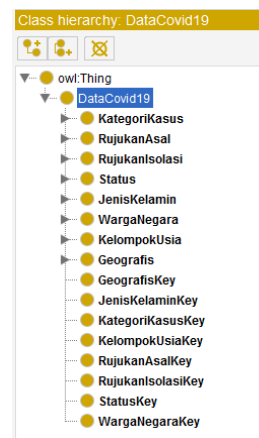


Figure 3. Classes of Covid-19 Incidences Data

- OWL Full For the use of maximum expressions with freedom in syntax, without any calculation
- Protégé

The Protege is made by an organization under the authorization of Stanford, which specializes in the field of ontology. Information relating to Protégé can be seen at the address [http:// Protege. stanford.edu/](http://Protege.stanford.edu/), including tutorials and the Protégé user community.

Protege is a tool used to create an ontology domain, adjust forms for data entry, and enter data. Supports various storage formats such as OWL, RDF, XML, and HTML. Protégé provides plug and play convenience that makes it flexible for developing prototype development. Protégé is created using the Java programming language. All tools in Protégé can be used via the Graphical User Interface

REFERENCES

1. Nicola Guarino, Pierdaniele Giaretta. Ontologies and Knowledge Bases Towards a Terminological Clarification
2. Noy, N.F., McGuinness D.L., Ontology Development 101 : Guide to Creating For Your First Ontology, Stanford Knowledge System Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880
3. A.Denicola et al.,2009. A Software Engineering Approach to Ontology Building, Information System 34(2).
4. *Maureen Donnelly; Giancarlo Guizzardi (2012)*. Formal Ontology in Information Systems: Proceedings of the Seventh International Conference (FOIS 2012). IOS Press. ISBN 978-1-61499-083-3.
5. Alani, H., TGVizTab: An Ontology Visualisation Extension for Protégé, Proceedings of Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering, Sanibel Island, Florida, USA, (2003)
6. J. Gennari et al. The evolution of Protégé: An environment for knowledge-based systems development. Int. Journal of Human-Computer Interaction, 58(1), 2003.
7. M. A. Musen et al. Component-based support for building knowledge-acquisition systems. In IFIP WCC Conference on Intelligent Information Processing, Beijing, China, 2000.
8. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, May, 2001.
9. Nugroho, A. (2012). Membangun Ontologi Jurnal Menggunakan Protege. Jurnal Transformatika,
10. [http:// Protege. Stanford.edu/](http://Protege.Stanford.edu/)