Semantic Representation for Health Data with Support to Spatial and Temporal Information

Dewa Rama[1] and Habibullah Akbar[2]

[1] **Computer Science Department, BINUS Graduate Program – Master of Computer Science**
Bina Nusantara University, Jakarta, Indonesia, 14080

[2] **Computer Science Department, BINUS Graduate Program – Master of Computer Science**
Bina Nusantara University, Jakarta, Indonesia, 14080

Abstract

Health data is one of information that has a close relationship with the spatial and temporal aspects. Spatial and temporal information has now become a more important component in health data analysis. This is based on the fact that there are an increasing aspect of geographical position and time constraints in public health data. Therefore, health data analysis can produce more meaningful information if the spatial and temporal information are well represented. The present model of spatial temporal representation generally offers model that satisfies only one form of representation, either spatial or temporal. This paper proposes a model of health data representation that supports the aspect of spatial temporal information using semantic web approach. The spatial aspect is represented using RCC8 topology relation and compass direction classification. While the temporal aspect is represented using Allen's temporal relation. Data representation is implemented in the form of ontology documents. The resulting ontology has DL expressivity of SHOIF(D), and has an average inference ratio of 15%.

Keywords: *health data representation, spatial and temporal representation, semantic web, ontology.*

1. Introduction

The semantic web provides the proper model in defining the meaning of objects, as well as linking them to the other related objects [1]. In semantic web technology, the formal meaning of the concept is expressively defined in one or more ontologies [2]. Ontology is a document that contains a collection of vocabulary, definitions, and relationships about a particular domain [3]. Ontology can explain the meaning of an object in a domain, and can also connect an object with its relation on other domains [4]. Health ontology contains formal meaning of human body anatomy, disease classification, symptom, treatment methods, type of medical personnel, as well as other things related to the field of health and medicine [5][6][7]. Because it stores the vocabulary, definitions, and relationships of a set of objects, ontology is then referred as the knowledge base in the semantic web technology [8].

In a semantic web environment, ontology documents are scattered across various service repositories in each of their respective domains [3]. Each ontology can connect with other ontologies, allowing objects and properties within them to connect to each other and to give a deeper meaning to each other [4]. For example, a climatic object on meteorological ontology may be connected to an object of disease on a health ontology. The relationship between the two objects can explain causality, factual statement, or other meanings. This meaning is then referred as the semantics, which is a more important concept of the relationship itself, which provides more valuable information for its users.

Spatial and temporal information has now become an important component in various types of applied applications [9]. This is based on the fact that there are aspects of geographical position and time constraints on some types of information available. Complex data structures become the challenge in managing information with spatial and temporal aspects, especially in the semantic web environment [10]. The solution required in managing spatial and temporal information, not only deals with how information is stored, but also how it can be accessed, while retaining the meaning of the actual facts.

The present model of spatial and temporal representation generally offers model that satisfies only one form of representation, either spatial or temporal. So far, SOWL is one of the appropriate model that offers some form of spatial and temporal representation into a single model. SOWL is a spatial and temporal representation that combines several common representation models, such as RCC8 topology relation and Allen's temporal relation [9]. The RCC8 topology relation is a spatial representation that supports the relationships between geographic positions. Whereas Allen's temporal relation is a temporal representation that supports the relative relationship between time intervals. Through such representation, SOWL allows spatial and temporal information to be represented in qualitative measurements.

2. Common Data Representations

2.1. Representation of Spatial Data

Spatial aspects are usually implemented in the form of two-dimensional (2D) or three-dimensional (3D) representation. In the Cartesian coordinate system, the 2D field involves two axes X and Y, while the 3D space involves the additional axis Z. In general, digital applications map the spatial aspect in the 2D field because of its easier implementation [11]. The relation between one location and another location is called a topology, a representation of relationships between geographic objects [12][13]. This concept not only determines the geometric location and shape of geographic objects, but also describes how they relate to each other, and how an area borders other neighboring regions [14].

2.1.1. RCC8 Topology Relation

Spatial information about the position of a geographic object is usually described in three ways, i.e. by topology, direction, and distance. Topology relations can be implemented using the RCC8 classification [9]. RCC8 is a formal form that classifies eight possible types of relationships from the relationship of two geographical objects.

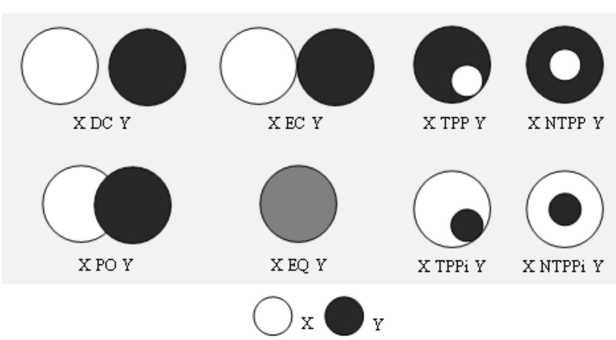

Fig 1. Spatial Classification of RCC8 Topology Relation [9]

Topological relationships explain the position of an object based on its relative topological position against other objects, such as "adjacent" or "opposite". Whereas the direction relation describes the position of an object by comparing it with the relative point of view of the other object, e.g. "on the north part" or "on the southeast side". Distance relation usually only explains quantitatively the position of an object with other objects through a certain amount of distance. Distance relation is usually combined with topology relation and direction relation to give better meaning.

2.1.2. Compass Direction Classification

The direction relation is generally implemented using the classification of the eight directions of the compass direction [9]. When implemented in a geographical area, the projection can be done by placing the position of an object in the coordinate system, then looking at it relative to the resulting compass direction division.

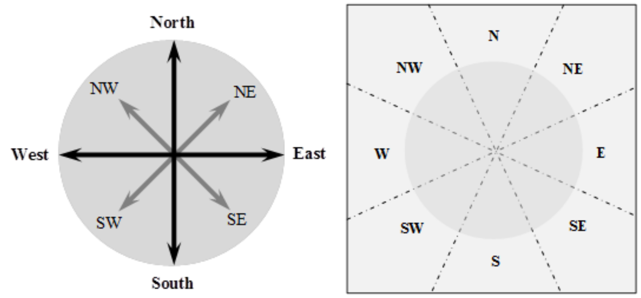

Fig 2. Spatial Classification of Compass Direction [9]

The relation of distance, as mentioned above, is quantitative and in its implementation combined with topology relation and direction relation. This allows the relation of distance to be qualitative, for example "X district is 25 km from the outer side of city Y" or "island A is located 100 km from north of island B". Ambiguous qualitative relationships, such as "near" or "far" can only be implemented if the qualitative measures are specified quantitatively. Specifications such as these can be specified on the application layer as well as on the data model layer.

2.2. Representation of Temporal Data

Time can be represented as discrete or continuous, linear or cyclical, absolute or relative, and quantitative or qualitative. Alternatively, time can also be represented using time points and time intervals [10]. On the one hand, there is a philosophical controversy about the change of objects over time, called the concept of perdurantist-endurantist, or the concept of beginning-ending [9]. This controversy relates to the question of the identity of an object when it is changed by time, especially when it forms a relation with another object. This concept ultimately indirectly affects the general representation of the spatial and temporal ontology models.

Relative time is usually represented both quantitatively and qualitatively. Quantitative representation can be done by specifying the value of a literal property with date data type. While qualitative representations are usually only specified in relative terms, such as "before" and "after". Time data types are fully supported in OWL, and have supported for time comparison operations, provided that each time point of the comparable object is known explicitly [15][16]. The probable problem is if only one of the objects specified the explicit time, where the other object specifies only the relative time to the object. This can only be solved with qualitative reasoning.

2.2.1. Allen's Temporal Relation

In operations and relations to time, there is a corresponding relationship between time point representation and time interval representation. The time point can be related to the "before", "after", and "simultaneously" operator [17]. These three relationships assume that each point of time lies in sequentially arranged timelines. Based on the sequential arrangement, the time interval can be directly represented as a pair of points on the sequential path. The time interval can be related to each relation operator at that point in time [9].

Allen's temporal relation is the relational operator used to operate the time qualitatively [10]. The After, OverlappedBy, MetBy, Contains, StartedBy, and FinishedBy relationships are the opposite of Before, Overlaps, Meets, During, Starts, and Finishes. A temporal relation can be one of the thirteen forms of Allen's temporal relation.

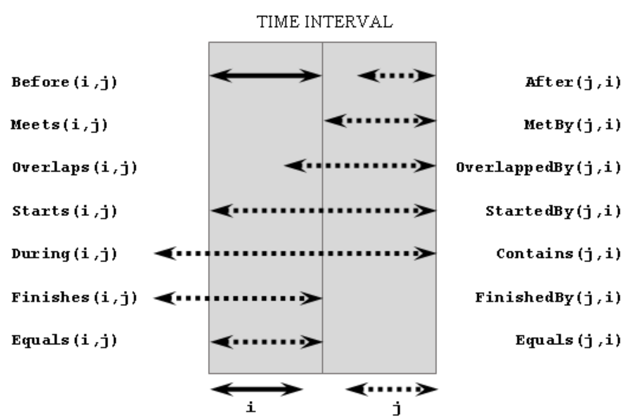

Fig 3. Allen's Temporal Relation [10]

Allen's temporal relation supports both time point and time interval qualitative relations. The qualitative relations can still be determined even though the explicit time is unknown, provided the relative time is known [10]. This of course increases the expression of time when applied to ontology.

3. Design of Semantic Data Representations

3.1. Semantic Representation of Health Data

Generic health data is defined as statistical information about a series of events and facts related to the health domain, such as disease spread, immunization distribution, health indicator, and so on. Thus, health facts are generally represented as instances of Event.

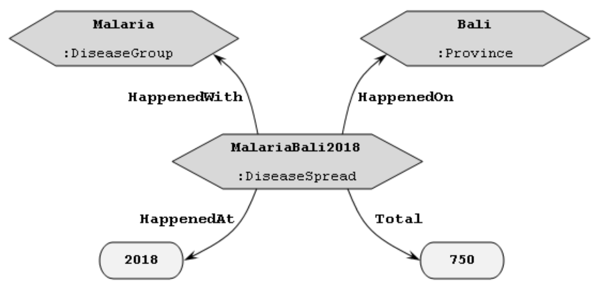

Fig 4. Representation of Health Data based on a Disease Fact

The illustration in Fig 4 describe a fact about the spread of Malaria disease in Bali that occurred during 2018. This fact is represented by MalariaBali2018, which is an instance of DiseaseSpread derived from the generic class Event. Furthermore, this object is associated with the Malaria object using the HappenedWith predicate, which states that this fact is an incident about malaria disease. The Malaria object is an instance of DiseaseGroup which defines the collection of disease.

Furthermore, the object is associated with Bali objects using the predicate HappenedOn, which states that this fact is an incident in the province of Bali. The object of Bali is an instance of class Province which is one of the classes in the spatial ontology. The object is then associated with a literal 2018 using the HappenedAt predicate, which states that this fact occurred in 2018. This literal value is only an integer, but in the next section it can be modeled with the temporal representation. Finally, the object is then connected with literal 750 through the Total predicate, which states that the facts occurred with the number 750 cases.

3.2. Semantic Representation of Spatial Data

Spatial data representation plays a role in providing the ability for a data in presenting facts about information relating to spatial aspects, such as maps, coordinates, and geographical position. Spatial data representation has been used extensively in various applications in geographic information systems, but its application is still minimal on semantic web technologies. This is due to the limited implementation of semantic web in research and geographic-based applications.

Spatial data representation will be designed by supporting some basic forms, such as latitude and longitude coordinates. In addition, to provide a more semantic meaning to spatial information, the data representation is complemented with the RCC8 topological relations and the compass direction classification [9]. By adding these two types of representation, any fact that has spatial information can be processed using a more complex spatial operation.

Spatial representation using the RCC8 topology relation allows machine to discover the broader semantic meaning of a spatial information. For example, Indonesia as a country encompassing Yogyakarta as a province, will form an NTPP topology relationship. Spatial ontology will then define that NTPP has an inverse on the NTTPi topology relation. Thus, the reasoner engine will soon understand that Yogyakarta is surrounded by Indonesia, which forms the NTPPi topology relation. This axiom can be obtained through reasoning even without explicitly defining it.

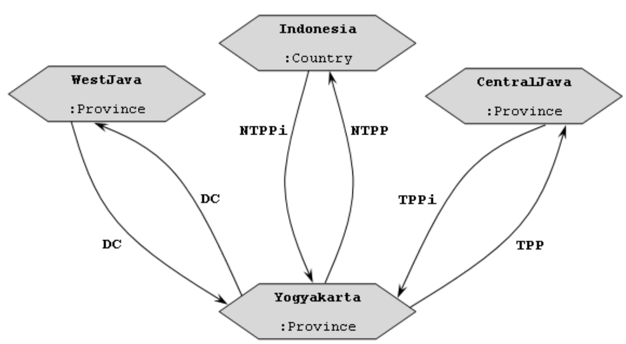

Fig 5. Spatial Representation using RCC8 Topology Relation

Furthermore, spatial representation using compass direction is also implemented to provide additional meaning about the position of an area directly to other areas. Although the compass direction is used generally to define the direct position, it also used to define the relative position. Relation represented by compass direction also can be recursive.

For example, if in the southeast of Bali is Australia, and in the southeast of Australia is New Zealand, then without explicitly defining, the reasoner engine can conclude that in the southeast of Bali there is also New Zealand.

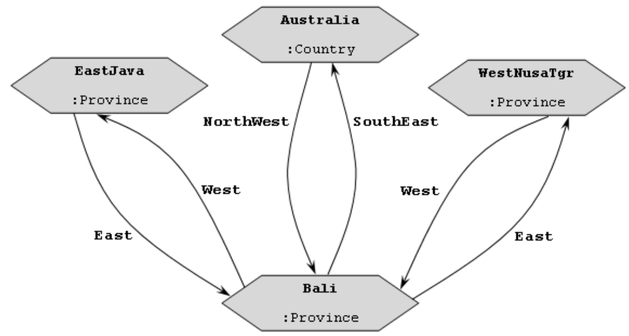

Fig 6. Spatial Representation using Compass Direction Classification

3.3. Semantic Representation of Temporal Data

Data that has time information is usually represented only by a primitive or literal data type. However, there is much other information that can be learned, or can be expressed by the reasoner engine, if the time information is represented not only through its value, but also through its meaning and relation to the time entity on other objects. Allen's temporal relation is an approach in time processing that supports more semantic time operations.

Allen's temporal relationships allow objects that have time information to be interconnected with each other. If there are two objects that have time information, but only one object defines its time value, then through Allen's temporal relation, the reasoner engine can still estimate the correct time position of the other object.

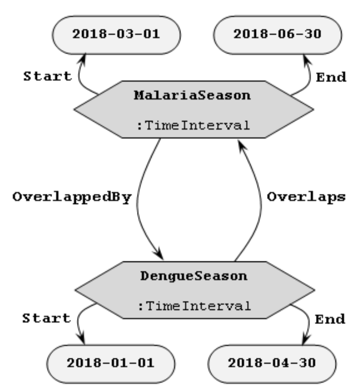

Fig 7. Time Representation using Allen's Temporal Relation

3.4. Representing Health Facts with Spatial and Temporal Aspects

A more complete semantic meaning will be generated when all representations are applied together. The illustration in Fig 8 shows the facts about the incidence of two types of diseases in different geographical location adjacent to the time interval. The dashed line represents automatically generated axioms obtained by the reasoner engine through inferencing the ontology.

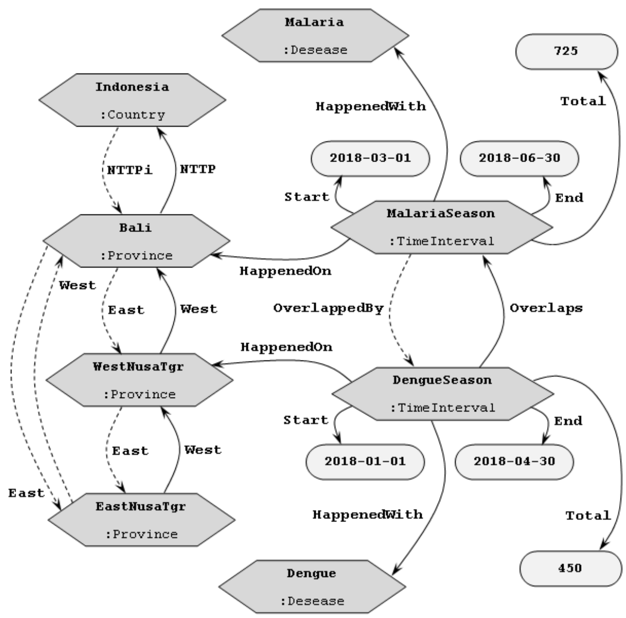

Fig 8. Representing Health Facts with Spatial and Temporal Aspects

4. Implementation

The data representation model is implemented into an ontology document, so it would be able to be used on semantic web technology stack. Health facts and information are presented through the instance of the Event class. These facts are linked by the HappenedWith object property. This object can be associated with spatial information, such as geographic location using the HappenedOn property. This object can also be associated with temporal information, such as the time of the event using the HappenedAt property. The number of facts of the event is represented using StatisticalAmout literal properties, which can be Total, Average, Mean, and so on.

Table 1. Semantic Specification of Health Data Representation

Class	Object Property	Literal Property
Event	HappenedWith, HappenedOn, HappenedAt	StatisticalAmout

Spatial information is represented using RCC8 topology and compass direction. Each type of relation is implemented as an object property with domain and range of AreaBoundary.

Table 2. Semantic Specification of Spatial Data Representation using RCC8 Topology Relation

Object Property	Inverse	Symmetric
Disconnected (DC)	-	Yes
Externally Connected (EC)	-	Yes
Equal (EQ)	-	Yes
Partially Overlapping (PO)	-	Yes
Tangential Proper Part (TPP)	TPPi	No
Tangential Proper Part Inverse (TPPi)	TPP	No
Non-Tangent Proper Part (NTPP)	NTPPi	No
Non-Tangent Proper Part Inverse (NTPPi)	NTPP	No

Table 3. Semantic Specification of Spatial Data Representation using Compass Direction Classification

Object Property	Inverse	Symmetric
North	South	No
NorthEast	SouthWest	No
East	West	No
SouthEast	NorthWest	No
South	North	No
SouthWest	NorthEast	No
West	East	No
NorthWest	SouthEast	No

The temporal information realized is represented using Allen's temporal relation. Each type of relation is implemented as an object property with a domain and range of TemporalRange, which can be either TimeUnit, or TimeInterval. TimeUnit is just a primitive data type of time, while TimeInterval is a class that specifically defines a time range consisting of StartTime and EndTime.

Table 4. Semantic Specification of Temporal Data Representation

Class	Object Property	Literal Property
TimeInterval	HappenedWith, HappenedOn	StartTime, EndTime, Total

Table 5. Semantic Specification of Allen's Temporal Relation

Object Property	Inverse	Symmetric
Before	After	No
After	Before	No
Meets	MetBy	No
MetBy	Meets	No
Overlaps	OverlappedBy	No
OverlappedBy	Overlaps	No
Starts	StartedBy	No
StartedBy	Starts	No
During	Contains	No
Contains	During	No
Finishes	FinishedBy	No
FinishedBy	Finishes	No
Equals	-	Yes

5. Evaluation

Evaluation in this study was conducted through a series of approaches to assess and measure the quality of the proposed data representation model. Evaluation of the data representation model is done through DL expressivity calculation and ontology reasoning experiment. These two evaluation methods together aim to measure the quality of the proposed data representation model for health data, as well as the support for spatial and temporal information.

5.1. DL Expressivity Calculation

The DL expressivity metric is a set of standard measures that express the degree of expressivity of an ontology specification [10]. DL expressivity metrics are calculated based on the number of axioms defined and related in an ontology [18]. If an ontology is related to another ontology, then DL expressivity metrics can be calculated individually or in whole. Evaluation of DL expressivity in this study was conducted on all ontologies produced, i.e. health, spatial, and temporal ontology. After that the measurement also done on the combination of all ontologies that have been integrated. The DL expressivity calculation is performed using the Protégé application and the DLMetric plugin.

Table 6. Axiom Statistics and DL Expressivity Metrics

Criteria	Value
All Axioms	1088
Logical Axioms	578
Declared Axioms	258
Class	41
Object Property	42
Data Property	23
Individual	155
Annotation Property	2
DL Expressivity	SHOIF(D)

Most highly expressive ontologies are satisfying the SOI fragments, such as SHOIN, SHOIQ, and SROIQ [19]. More specifically, SOI fragments are met with the application of axioms of attributive facts, atomic, sliced, enumerated, negated, transitive, nominal, and inverse. The calculated expressivity in the ontology generated from this study has also fulfilled the specification of the HF(D) fragment, i.e. the axioms of class hierarchy, function, and the type of formation data.

5.2. Ontology Reasoning Experiment

An expressive ontology should be able to find new implicit meaning that can be learned by the reasoner engine. The reasoner engine works by learning in all defined axioms in an ontology. Furthermore, the reasoner engine will infer some new facts and explore new knowledge that has not explicitly defined in ontology.

The ontology reasoning evaluation in this study was performed on each ontology generated, then continued on the compilation of all the ontologies. The evaluation process was carried out using the HermiT reasoner engine. Table 7 presents the results obtained when the reasoning process is complete.

Table 7. Results of Ontology Reasoning Experiment

Ontology	Existing Axioms	Inferred Axioms	Inference Ratio
Health	336	60	17%
Spatial	343	82	23%
Temporal	399	41	10%
Compiled	1088	173	15%

From the analysis results, it appears that spatial ontology has the highest inference ratio, while the temporal ontology has the lowest inference ratio. In addition, the yield of inference ratio to the existing axioms was 15% on average.

6. Conclusion and Future Work

The proposed data representation can be implemented in public health data domain, which has extensive aspects in spatial and temporal information. The data representation can present semantic meaning and operation to spatial and temporal information. The resulting ontology has a DL expressivity metric of SHOIF(D), which can be classified as a high-expression ontology. Through the learning process by the reasoner engine, the resulting ontology has an average inference ratio of 15%.

The data representation model proposed in this study can be piloted in other cases in addition to health data. The spatial representation still can be improved to support related information, such as weather. The time information on the proposed temporal representation can be further developed by providing support to information history.

References

[1] Haendel, M. Neuhaus, F. Sutherland, D. Mabee, P. Mejino, J. Mungall, C. Smith, B. (2007) CARO: The Common Anatomy Reference Ontology. Anatomy Ontologies for Bioinformatics, pp 327-349.

[2] Keskes, N. Rahmoun, A. (2015) Meaning Negotiation Based on Merged Individual Context Ontology and Part of Semantic Web Ontology. International Journal on Human Machine Interaction. Vol. 2, No. 1.

[3] Vandenbussche, P. Atemezing, G. (2014) Linked Open Vocabularies (LOV): A Gateway to Reusable Semantic

Vocabularies on the Web. Journal of Semantic Web. Vol. 1, pp. 1–5.

[4] Le, D. Heon, M. Volmer, N. (2017) Towards Semantic Applications: From Knowledge Management to Data Publishing on the Web. 2017 International World Wide Web Conference Committee. Perth, Australia.

[5] Diaz, M. Antonelli, L. Sanchez, L. (2017) Health Ontology and Information Systems: A Systematic Review. IEEE Latin America Transactions. Vol. 15, No. 1.

[6] Sen, P. Banerjee, S. Mukherjee, N. (2017) Ontology Driven Approach to Health Data Management for Remote Healthcare Delivery. Mobile Health Proceedings of the 7th ACM Workshop on ACM. Article No. 2.

[7] Sepulveda, A. Chavez, E. Motz, R. (2017) An Ontology to Assess Data Quality Domains: A Case Study Applied to Healthcare Entity. IEEE Latin America Transactions. Vol. 15, No. 8, pp. 1506-1512.

[8] Bennett, M. Baclawski, K. (2017) The Role of Ontologies in Linked Data, Big Data and Semantic Web Applications. Journal of Applied Ontology, Vol. 12, pp. 189-194.

[9] Batsakis, S. Petrakis, E. (2010) SOWL: Representing Spatio Temporal Information in OWL. The 7th Extended Semantic Web Conference. Poster Paper.

[10] Zamborlini, V. Guizzardi, G. (2010) On the Representation of Temporally Changing Information in OWL. In Enterprise Distributed Object Computing Conference Workshops, pp 283-292.

[11] Katz, Y. Grau, B. (2005) Representing Qualitative Spatial Information in OWL-DL. The OWL Experiences and Directions CEUR Workshop Proceedings. Vol. 188, No. 1.

[12] Hofmeyer, H. Schevenels, M. Boonstra, S. (2017) The Generation of Hierarchic Structures Via Robust 3D Topology Optimization. Elsevier Journal in Advanced Engineering Informatics. Vol. 33, pp. 440-455.

[13] Kim, K. Carlis, J. Keefe, D. (2017) Comparison Techniques Utilized in Spatial 3D and 4D Data Visualizations: A Survey and Future Directions. Elsevier Journal in Computers and Graphics. Vol. 67, pp. 138-147.

[14] Sotnykova, A. Vangenot, C. Cullot, N. Bennacer, N. Aufaure, M. (2005) Semantic Mappings in Description Logics for Spatio Temporal Database Schema Integration. Journal on Data Semantics, pp 143-167.

[15] Soylu, A. Giese, M. Schlatte, R. Ruiz, E. Kharlamov, E. Ozcep, O. Neuenstadt, C. Brandt, S. (2017) Querying Industrial Stream Temporal Data: An Ontology Based Visual Approach. Journal of Ambient Intelligence and Smart Environments. Vol. 9, No. 1, pp. 77-95.

[16] Brandt, S. Kalayci, E. Ryzhikov, V. Xiao, G. Zakharyaschev, M. (2017) A Framework for Temporal Ontology Based Data Access: A Proposal. Journal of Advances in Databases and Information Systems. pp. 161-173.

[17] Krieger, H. (2010) General Methodology for Equipping Ontologies with Time. The 7th International Conference on Language Resources and Evaluation, ELRA, pp 3165-3172.

[18] Buhmann, L. Lehmann, J. Westphal, P. Bin, S. (2018) DL Learner: Structured Machine Learning on Semantic Web Data. Companion Proceedings of the Web Conference, pp 467-471.

[19] Horridge, M. Aranguren, M. Mortensen, J. Musen, M. Noy, N. (2012) Ontology Design Pattern Language Expressivity Requirements. WOP Proceedings of the 3rd International Conference on Ontology Patterns. Vol. 929, pp 25-36.

Dewa Rama is a student for master degree in informatics engineering from Bina Nusantara University. Major in software engineering, information retrieval, and knowledge management.

Habibullah Akbar is a post graduate lecturer in Bina Nusantara University. Major in software engineering, data mining, machine learning, information retrieval, computer vision, and business intelligence.