Solving the Complexity of Heterogeneity Data on Learning Environment Using Ontology

Arda Yunianta*[1,2], Mohd Shahizan Othman*[1], Norazah Yusof*[1,3], Lizawati Mi Yusuf[1], Juwairiah[4], Nurul Syazana Selamat[5]
[1] Faculty of Computing, Universiti Teknologi Malaysia,81310, Malaysia
[2] Faculty of Information Technology and Communication. Mulawarman University, 75119, Indonesia
[3] Faculty of Computing and Information Technology, King Abdul aziz University, 21911, Saudi Arabia
[4] Faculty of Industrial Technology. Universitas Pembangunan Nasioanal Veteran, 55283, Indonesia
[5] Centre for Information and Communication Technology, University Teknologi Malaysia,81310, Malaysia
*Corresponding author, e-mail: arda.mti08@mail.ugm.ac.id; norazah64@gmail.com; shahizan@utm.my

Abstract

Distributed and various systems on learning environment are the current issues to produce big data and heterogeneity data problem. Heterogeneity on learning environment is about numerous learning applications and various learning information to support a learning process in educational institutions. There are a lot of relationships are formed between elements on learning environment. The elements on learning environment consist of learning data, learning applications, data sources, learning concept, and data heterogeneity aspect on learning environment. These elements are interrelated and produce complex relationship between each other. A complex relationship problem between elements on learning environment makes a process of analysis and identification difficult to be done. Existing method to drawing this heterogeneity problem make confuse and misunderstanding readers. To solved this problem, researcher using ontology knowledge to describe and draw a semantic relationship that represent the complexity of data relationship on learning environment. The result of this analysis is to develop ontology knowledge to solve heterogeneity data problem specific in complexity relationship on learning environment. This result can give better understanding to the readers about complex relationship between elements on learning environment.

Keywords: learning environment, data heterogeneity, ontology knowledge, semantic approach

1. Introduction

Implementation of Electronic system on learning environments is becoming popular and very important in today's scenario because of their flexibility, convenience and accessibility to support learning activities in traditional learning process [1],[2]. There is numerous and various application systems on learning environment from different function and with specific purpose, this is usually known as heterogeneity on learning environment. The heterogeneity may be the difference in: User interface, Platform, Application system, Database system, Data representation etc.

The heterogeneity of data is a current issue in distributed and various information sources. Development of applications and information systems makes heterogeneity problems grow up and more complex, and from that problems need to find the best solution [3],[4]. Data on learning environment is increasingly grown up and becoming more meaningful to support learning activities [5],[6].

Heterogeneity of data on learning environment is about different data representation and types of information or data in different and numerous applications to support a learning process in education institutions [7]. Different applications are develop for specific purposes based on function and feature that included on that applications [5]. A lot of applications developed on learning environment, such as Teaching and learning online application, Library application system, Question bank system, Student management and payment system, Academic information management system, Student registration system and subject course evaluation system. In this paper, researchers are using UTM (Universiti Teknologi Malaysia) as a case study to analyze the data heterogeneity problem on university environment. With numerous applications that develop with various system and database schema, produces a big data with heterogeneity problem on that environment.

The aim in this paper is to analyze and to identify what kind of data heterogeneity problem that be happen on learning environment, what are the semantic relationship relate between elements on data heterogeneity on learning environment. After analyzing and identification process, the main contribution of this research is to develop ontology knowledge to describe what kind of data heterogeneity problem on learning environment and to get better reader understanding with ontology viewing that contain semantic relationship between elements on learning environment.

We describe that there are four main aspects of heterogeneity on learning environment. The four aspects are heterogeneity data, learning resources, applications, and learning theories. Details heterogeneity criteria's will discuss in the next section.

This paper is extended paper from this paper [20]. To complete the research, this paper divided into several stages, the first is to analyze and identify four main aspects of heterogeneity on learning environment. The next step is to create semantic relationship between all of elements on learning environment. And the final result is to develop ontology knowledge to draw the complexity that happen on data heterogeneity problem on learning environment with semantic relationship.

2. Heterogeneity Aspects On Learning Environment
2.1. Heterogeneity on Data

In the heterogeneity of data aspect we describe that there are five sub aspects with some of the elements contained in it. The five sub aspects are data types, data structure, learning information, database system, and data representation. In the next paragraph we will explain more detailed about this five sub aspects.

There are numerous data type format and data structure format [8],[9]. For the data structure format there are three types of **data structure**, there are unstructured data, semi-structure data and structured data. Different with data structure, **data types** have five data types, they are multimedia data, images data, text data, web/web-service data and database data. With numerous applications that develop with various system and database schema, produces a big data with heterogeneity problem on that environment. Different application system with numerous and heterogeneity information, data sources, databases system and data representation makes communication and integration process between this applications difficult to implemented [8],[10].

Database system aspect also completed the data heterogeneity problem on learning environment. Some applications are develop under the same database system and the other applications are developed with different database system. Researchers are identifying a several database system to develop an application, there are Oracle database system, MySQL database system, SQL Server, and Postgre SQL. Different data are saved in the database system also have different schemas to develop the database system this is produce a different data representation format.

Development and implementation system on learning environment produce a differences in **data representation** aspect. System development with different developer also makes differences in data representation schema on database system.

Heterogeneity data specific on **learning information** will have interrelation with constructive alignment theory that have three element with one extra element. The three elements there are teaching and learning Experiences, assessment tasks and learning outcomes. Whereas, the extra element is relate to the purpose of constructive alignment theory, namely student result.

2.2. Heterogeneity on Learning Resources

Learning resources is relate with reference sources as a references for students to get better learning knowledge and to finishing all assignment that given by lecturers. In the learning process conduct by students and lecturers they have reference sources. The reference sources is not only from lecturer slides but there are a lot of learning material from the other sources such as journal papers, website pages, books, articles, Instagram, YouTube and the other external sources [11]. From these sources students can learn not just from text book, but

students also can learn from audio and video content on internet. These learning sources relate with data types and data structures element on the other heterogeneity aspects.

2.3. Heterogeneity on Applications

Nowadays, application system development on learning environment is growing fast. A lot of application system developed to help learning process in some education institutions. Every application is develop for specific purposes with several facilities. Each application is developing with different function, purpose and with different developer produce heterogeneity aspect on the application perspective. Application developer under different programming language, system architecture, model, interface, platform and databases system make communication and integration process more difficult.

Data heterogeneity on applications aspect is about different data representation and types of information or data in different and numerous applications to support a learning process in education institutions [7]. Different applications are develop for specific purposes based on function and feature that included on that applications [5]. A lot of applications developed on learning environment, such as teaching and learning online application called Moodle e-learning, academic information management system, student management and payment system, Question Bank System, Subject Courses evaluation system, student registration system, library application system and other learning applications. In this research, researchers are using UTM (Universiti Teknologi Malaysia) learning applications as a case study to analyze the data heterogeneity problem on applications.

2.4. Heterogeneity on Learning Theories

Learning theories aspect is relate with learning information and the other elements on learning environment. There are two learning theories relate on this research there are constructive alignment and intelligent tutoring system.

Constructive alignment is an educational model to conduct learning process. Two aspect are built on constructive alignment is a constructive aspect refers to the idea that students construct the meaning on learning process through relevant learning activities.The key is that the components in the teaching system, especially the teaching methods used and the assessment tasks are aligned with the learning activities assumed in the intended outcomes. The learner is a peoples who want get better understanding to solve learning problem that given to them [12],[13].

To conduct a better learning process trough constructive alignment is how to perform all assessment tasks and teaching learning process (included content and methods) must be linked to the desired unit of study learning outcomes. Figure 1 shows the interrelationship between the three elements on constructive alignment concept [14].
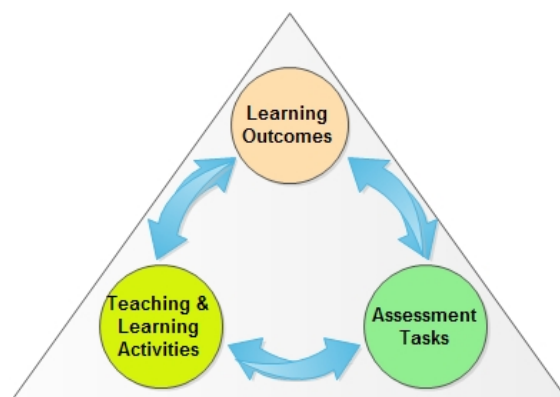

Figure 1. Constructive Alignment Theory, adapted from [13]

The important thing on a constructive alignment is to get balance between teaching and learning activities, assessment tasks activities and learning outcomes to be achieved. A Constructive alignment is an educational model [15], this concept based upon constructive learning theory (constructivism) and aligned curriculum. The aim of this concept is get better students' performance to enhance student outcomes. While the model has been implemented on many domains and linked to positive effect to the student results, also for learning environment this concept really closed with all of activities on learning process [16].

Constructive Alignment concept is comprised by a set of principles that can be used to devising Teaching and Learning Activities that help in achieving the Intended Learning Outcomes. This is accomplished by carefully aligning and learning assessments to support the students to fulfill the Intended Learning Outcomes [12]-[14].

Learning model is a part of **Intelligent Tutoring Systems** concept that completed with interface as a bridge to communicate with user learner. Intelligent tutoring systems contain four components: domain, student, tutor and user interface [16]. Intelligent tutoring systems (ITSs) must be equipped with an explicit representation of the domain knowledge that is the subject of the learning activity to enable learners to acquire knowledge and develop skills in a specific domain. It must also be equipped with the mechanisms by which can be to solve problems in the domain through acquired knowledge and better skill development, because this is a main purpose of intelligent tutoring systems [17]. The purpose of the learning process is how to provide better personalized service and teaching materials for learners to achieved better learners result. Figure 2 shows the interrelationship three elements on learning model as a main part of intelligent tutoring system [17].
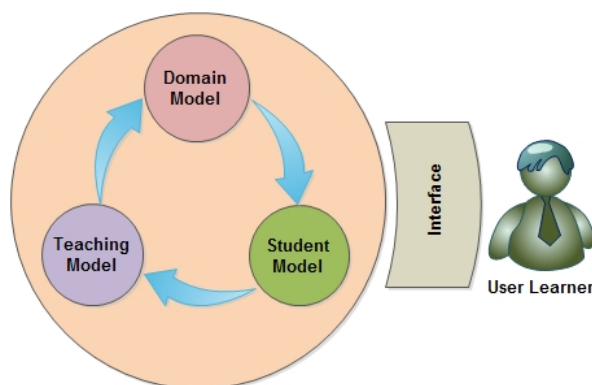

Figure 2. Intelligent Tutoring Systems, adapted from [17]

The **domain model** is also called expert knowledge. This model contains the problem solving strategies, rules and concept of the domain to be learned. It can fulfill several roles: as a source of expert knowledge, a standard for evaluating the student's performance or for detecting errors, etc. The domain model can organized to be a curriculum, learning structure that including all the learning knowledge elements and linked together according to pedagogical sequences [16]. The **student model** is the core component of an ITS. Ideally, it should contain as much knowledge as possible about the students cognitive that support learning process to get better understanding of student's perspective to solve learning problem to get better students result [16]. The **teaching model** receives input from the domain and student models and makes decisions about tutoring strategies and actions. Based on principled knowledge, it must make such decisions as whether or not to intervene, and if so, when and how. Content and delivery planning are also part of the tutoring model's functions. Tutoring decisions would ideally be reflected in different forms of interaction with the student: Socratic dialogs, hints, feedback from the system, etc. More generally, student and teaching interactions usually occur through the learning **interface**, also known as the communication or interface component. This component gives access to the domain knowledge elements through multiple forms to interact with user learner.

3. Semantic Relationships Between Elements On Learning Environment

After analysis and identification process, the next step is to draw the complexity relationship between elements on learning environment. Table 1 shows the complexity in heterogeneity aspect on learning environment and shows the types of realtionships between elements.

Table 1 shows a lot of relationship between elements on learning environment. From this table can be seen and conclude that it is very complex the relationship between elements. The existing method proves that the representation of heterogeneity on learning environment still have a weaknesses and still need to improve.

Table 1. Learning Information Semantic Relationship

Elements	Type of Relationship	Elements	Elements	Type of Relationship	Elements
Websites	hasATypeOfData	Web/Web-Service Data	Youtube	hasADataStructure	Unstructured Data
	hasATypeOfData	Database Data		hasATypeOfData	Multimedia Data
	hasADataStructure	Structured Data	Books	hasATypeOfData	Text Data
	hasADataStructure	Semi Structured data		hasADataStructure	Unstructured Data
Journal Paper	hasATypeOfData	Text Data	Instagram	hasADataStructure	Unstructured Data
	hasATypeOfData	Web/Web-Service Data		hasADataStructure	Semi Structure Data
	hasADataStructure	Unstructured Data		hasATypeOfData	Multimedia Data
	hasADataStructure	Semi Structured data		hasATypeOfData	Text Data
Learning Outcomes	CategorizeAs	Domain Model		hasATypeOfData	Web/Web-Service Data
	CategorizeAs	Teaching Model	Assessment Task	IsA	Learning Information
	CategorizeAs	Student Model		CategorizeAs	Domain Model
	IsA	Learning Information		CategorizeAs	Teaching Model
Teaching and Learning Online Application	HasA DataRepresentation	Data Representation 5 in Application 5	Student Result	IsThePurposeOf	Student Model
				IsThePurposeOf	User Interface
			Academic Information Management System	HasA DataRepresentation	Data Representation 2 in Application 2
	HasADatabaseSystem	MySQL		HasADatabaseSystem	Oracle
	HasA	User Interface		HasA	User Interface
	CategorizeAs	Student Model		CategorizeAs	Student Model
	CategorizeAs	Teaching Model		CategorizeAs	Domain Model
Question Bank System	HasA DataRepresentation	Data Representation 4 in Application 4	Student Management and Payment System	HasA DataRepresentation	Data Representation 1 in Application 1
	HasADatabaseSystem	Oracle			
	HasA	User Interface		HasADatabaseSystem	MySQL
	CategorizeAs	Teaching Model		HasA	User Interface
	CategorizeAs	Domain Model		CategorizeAs	Student Model
Subject Course Evaluation System	HasA DataRepresentation	Data Representation 3 in Application 3		CategorizeAs	Domain Model
			Library Application System	HasA DataRepresentation	Data Representation 6 in Application 6
	HasADatabaseSystem	SQL Server		HasADatabaseSystem	Postgre SQL
	HasA	User Interface		CategorizeAs	Student Model
	CategorizeAs	Teaching Model		HasA	User Interface
Student Registration System	HasA DataRepresentation	Data Representation 7 in Application 7	Teaching and Learning Experiences	CategorizeAs	Domain Model
	HasADatabaseSystem	MySQL		CategorizeAs	Teaching Model
	CategorizeAs	Student Model		CategorizeAs	Student Model
	HasA	User Interface		IsA	Learning Information
Student Result	IsThePurposeOf	Teaching and Learning Experiences	Student Result	IsA	Learning Information
	IsThePurposeOf	Domain Model		IsThePurposeOf	Learning Outcomes
	IsThePurposeOf	Teaching Model		IsThePurposeOf	Assessment Task

Every element on learning environment at least has one relationship, and some elements on learning environment can have a lot of relationships. This situation makes difficult to get clear view and better understanding, because there are no description name of relationship between elements on learning environment. Even though there are name of description of relation between these elements still difficult to get understanding and to analyze the picture. **Ontology gives a better solution** to solve this problem because ontology gives a semantic relationship attribute on every element on learning environment [19].

From Table 1 we conclude that there are 68 relationships between elements on learning environment with 8 types of relationship, there are *hasATypeOfData, hasADataStructure, hasADatabaseSystem, hasADataRepresentation, hasA, categorizeAs, isA, isThePurposeOf*.

Table 1 shows three columns with two columns are elements that have interrelationship between each other and one column is type of relationship of elements on learning environment. From this type of relationship we can build ontology knowledge to represent the heterogeneity on learning environment to give a clear image and better understanding about the complexity of heterogeneity on learning environment.

4. Ontology Knowledge Data Heterogeneity On Learning Environment

Analysis and identification process to describe a complexity on data heterogeneity on learning environment have shown on previous chapter on this paper. After create a semantic relationship between elements on learning environment, the next step is to developing ontology to shows more derailed all of semantic relationship on data heterogeneity on learning environment.

Data heterogeneity complexity shows on Table 1 as an explanation to shows all of elements on data heterogeneity on learning environment and relationship between elements. After this, researcher will use ontology approach to get better representation from complexity relationship between elements on learning environment. Before develop semantic relationship between elements.
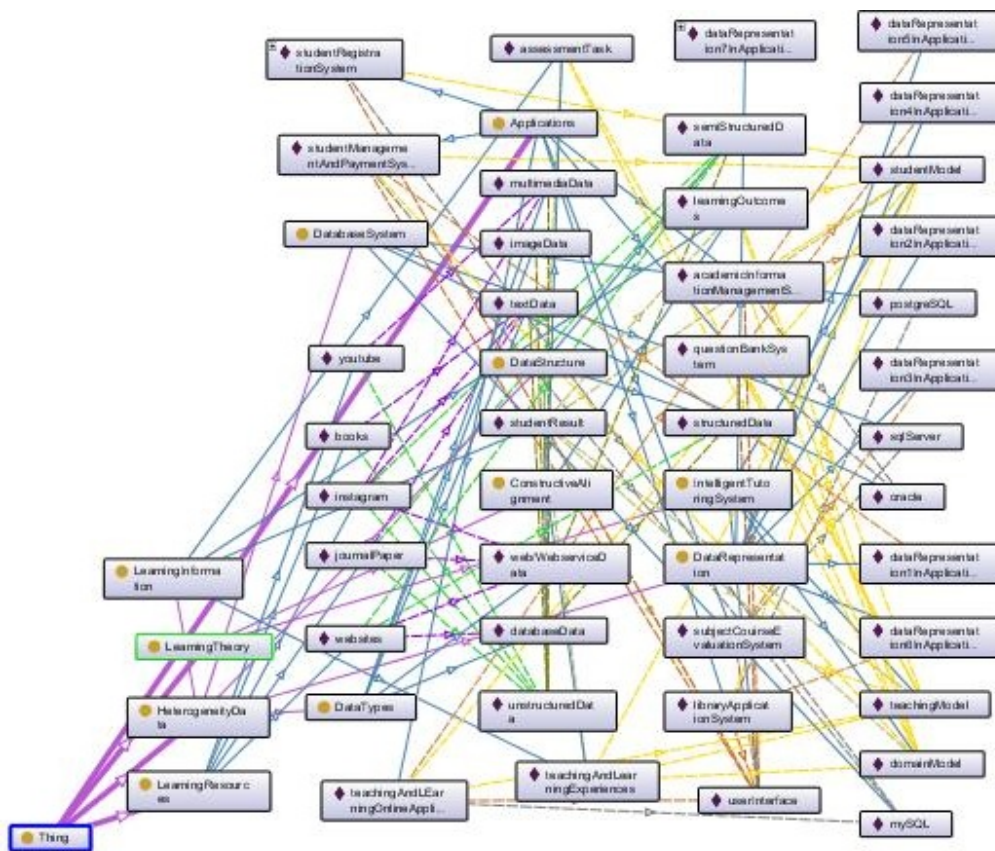

Figure 3. Ontology of Heterogeneity aspects on Learning Environment

Ontology knowledge that shows on Figure 3 is complete ontology knowledge with semantic relationship between all of elements on learning environment. **Applications class** has seven instances, there are *student Management And Payment System, student Registration System, question Bank System, library Application System, subject Course Evaluation System, academic Information Management System, and teaching And Learning Online Application*. **Learning Resources** class that have youtube, books, Instagram, journalPaper and websites. **Learning Theory** class with two subclass and seven instances. In the *Intelligent Tutoring System* subclass has *teaching Model, domain Model, student Model* and *user Interface*. Where as for *Constructive Allignment* subclass has *teachingAndLearning Experiences, learning outcomes and assessment Task*. **Heterogeneity Data** class that has five subclasses and twenty three instances. In the learning Information subclass has assessment Task, learning Outcomes, teaching And Learning Experiences and student Result. In the Data Structure subclass have semi Structured Data, structured Data and unstructured Data. In the Database System subclass have oracle, MySQL, sql Server and postgre SQL. In the Data Types subclass have multimedia Data, images Data, text Data, web/Web Service Data and database Data. And the last subclass is the Data Representation subclass has data Representation 1 In Application 1, data Representation 2 In Application 2, data Representation 3 In Application 3, data Representation 4 In Application 4, data Representation 5 In Application 5, data Representation 6 In Application 6 and data Representation 1 In Application 7.

Every class, subclass and instance on ontology possible to have more than one relationship. On this ontology knowledge researcher describe a four main class under root class (Thing), there are learning information, heterogeneity data, learning model and constructive alignment class. From four classes there are two classes that have subclass as a detail of each class. Heterogeneity class has five sub classes are database system, data representation, data structure, data types and learning knowledge.

In the ontology knowledge sometimes they have same instance in different class or subclass. This occurs in constructiveAllignment subclass and learningInformation subclass, there are have three same instances with one extra instance on learningInformation subclass. This is because of in these two subclasses have three same instances that have interrelationship between them.

5. Conclusion

A complexity relationship problem on learning environment is very difficult to draw and describe using basic method. There are many things on heterogeneity aspect on learning environment such as elements, aspects and relationship between them. The element on learning environment consists of learning model, learning applications, learning sources, and learning data concept (constructive alignment). While for aspects on learning environment is a data heterogeneity aspect on learning environment. Data heterogeneity aspect on learning environment consists of learning knowledge, data types, data structure, database system and data representation. All of elements and aspects on learning environment are interrelated and produce complex relationship between each other. A complex relationship problem between elements on learning environment makes a process of analysis and identification difficult to be done. Semantic technology through ontology knowledge is a current approach to solve a complexity relationship on some domain. Ontology knowledge using a semantic relationship is to handle a numerous and various relationship that exists on that. Ontology knowledge is a better solution to handle data heterogeneity problem on learning environment.

References

[1] Dietinger T. *Aspects Of E-Learning Environments.* Austria. Graz University of Technology. 2003.
[2] Xiaofei L, Saddik AE, Georganas ND. *An implementable architecture of an e-learning system.* Electrical and Computer Engineering, 2003 IEEE CCECE 2003 Canadian Conference on 2003; 2: 717-720.
[3] Kim W, Seo J. Classifying schematic and data heterogeneity in multi database systems. *IEEE Computer.* 1991; 24: 12-18.
[4] Kashyap V, Sheth A. *Semantic heterogeneity in global information systems: The role of metedata, context and ontologies.* Papazoglou MP, Schlageter G, editors. Cooperative information systems. San Diego. Academic Press. 1997: 139–178.
[5] Shyamala R, Sunitha R, Aghila G. Towards Learner Model Sharing Among Heterogeneous E-Learning Environments. *International Journal of Engineering Science and Technology (IJEST).* 2011; 3: 2034-2040.
[6] Qomaruddin M, Rahman AA, A Lahad N. CAPBLAT: An Innovative Computer-Assisted Assessment Tool for Problem Based Learning. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2014; 12(1): 241-250.
[7] Yunianta A, Yusof N, Jayadianti H, Othman M, Suhaimi S. *Ontology Development to Handle Semantic Relationship between Moodle E-learning and Question Bank System.* In: Herawan T, Ghazali R, Deris MM, editors. Recent Advances on Soft Computing and Data Mining: Springer International Publishing. 2014: 691-701.
[8] Dong MH, Yan LD, Ming HZ, Chi Z. *Application of ontology-based automatic ETL in marine data integration.* Electrical & Electronics Engineering (EEESYM), 2012 IEEE Symposium on. 2012: 11-23.
[9] Zheng L, Terpenny J. A hybrid ontology approach for integration of obsolescence information. *Computers & Industrial Engineering.* 2013; 65: 485–499.
[10] Wang CC, Pai WC, Yen NY. *A Sharable e-Learning Platform Based on Cloud Computing.* 3rd International Conference on Computer Research and Development (ICCRD*).* 2011: 1-5.
[11] Gudanescu N. Using modern technology for improving learning process at different educational levels. *Procedia - Social and Behavioral Sciences.* 2010; 2: 5641-5645.
[12] Biggs J. *Aligning teaching for constructing learning.* Buckingham. Open University Press. 2003.
[13] Biggs J, Tang C. *Teaching for Quality Learning at University.* Berkshire, England. Open University Press. 2007.
[14] Biggs J. Enhancing teaching through constructive alignment. *High Educ.* 1996; 32: 347-364.
[15] Cain A, Woodward CJ. *Toward constructive alignment with portfolio assessment for introductory programming.* IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE). 2012: 11-17.
[16] Nkambou R, Bourdeau J, *Mizoguchi R. Introduction: What Are Intelligent Tutoring Systems, and Why This Book?* In: Nkambou R, Bourdeau J, Mizoguchi R, editors. Advances in Intelligent Tutoring Systems: Springer Berlin Heidelberg. 2010: 1-12.
[17] Nkambou R. *Modeling the Domain: An Introduction to the Expert Module.* In: Nkambou R, Bourdeau J, Mizoguchi R, editors. Advances in Intelligent Tutoring Systems: Springer Berlin Heidelberg. 2010: 15-32.
[18] You D, Shen L, Peng S, Liu J. *Flexible Collaborative Learning Model in E-Learning with Personalized Teaching Materials.* In: Jin D, Lin S, editors. Advances in Computer Science, Intelligent System and Environment: Springer Berlin Heidelberg. 2011: 127-131.
[19] Firdausiah AB, S. DO, Yuhana UL, Kita T. Sistem Penilaian Otomatis Jawaban Essay Menggunakan Ontologi Pada Moodle. *TELKOMNIKA Telecommunication Computing Electronics and Control.* 2008; 6(3): 167-175.
[20] Yunianta A, Yusof N, Othman MS, Aziz A, Dengen N. *Analysis and Identification of Data Heterogeneity on Learning Environment Using Ontology Knowledge.* International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2014). Yogyakarta, Indonesia: IAES. 2014: 154-160.