Weighted Ontology and Weighted Tree Similarity Algorithm for Diagnosing Diabetes Mellitus

Widhy Hayuhardhika Nugraha Putra
Program of Informatics and Computer Science
Universitas Brawijaya
Malang, Indonesia
widhy@ub.ac.id

Sugiyanto
Department of Informatics
Institut Teknologi Adhi Tama Surabaya
Surabaya, Indonesia
sugiyanto.lecture@gmail.com

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Mohamad Sidiq
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
sidiq12@mhs.if.its.ac.id

***Abstract*— Application knowledge base for diabetes such as expert systems has been developed, but generally using conventional methods that have limitations in representing knowledge. Ontology supports the search of data / information by defining the concept of convergent intended by the user. This study using Diabetes Mellitus Classification based diabetes disease diagnosis from World Health Organization Geneva. This system receives input patient data from user. Then, system will build the patient ontology to represent patient knowledge. We are connecting Java applications to Protégé using OWL API. Then, system will calculate the weight of an ontology based on density. This system use JENA Inference Engine and working memory area for reasoning. The system would then do process similarity matching with Ontology Diabetes Mellitus using weighted tree similarity algorithm. Ontology has the highest similarity value will be the proposed diagnosis. Results of this study show that the representation in the form of OWL ontology using weighted ontology and weighted tree similarity algorithm can be used to represent knowledge about diabetes mellitus.**

Keywords—diabetes mellitus; semantic search; weighted ontology; weighted tree similarity

I. INTRODUCTION

Diabetes mellitus is a disease in which the pancreas cannot produces enough insulin. Insulin is a hormone that allows glucose (sugar) to absorbed into the cells of the body. The body cannot use glucose properly, that glucose levels in the blood is high. This condition causes damage important organs, and serious health complications as stroke, blindness, heart disease, and renal failure. Therefore, diabetes must be controlled.

People with diabetes mellitus are five times more potential than people who not has diabetes to had a heart attack. In fact, 70 percent of people with diabetes mellitus die from stroke or heart attack. Application knowledge base for diabetes such as expert systems has been developed, but generally using conventional methods, that have limitations in representing knowledge [1].

The meaning and content of the domain knowledge using a relational database or as structured documents have weaknesses such as eliminating many semantic information, fewer supportive of the search process knowledge from a variety of user perceptions, and even make it difficult to develop the model along with the development of knowledge. One’s of the main constraints of the relational data model approach is maintaining consistency, drop redundancies, and drop anomaly is done by establishing model data relationships in a normalized table [2].

Ontology technology is new ideas in knowledge representation techniques are much more expressive than conventional knowledge representation techniques. In addition it also has the advantage of ontology-based semantic query than traditional techniques based syntactic query. Ontology further used in information retrieval as search strings in web search engines, and knowledge management [3]. The domain ontology framework that must made clear, to able to take right information.

In semantic search using weighted tree similarity algorithm, metadata is based on labeled tree with nodes, branches labeled and weighted. Metadata tree structure is based on a kind of semantic information taxonomy, ontology, preference, synonyms, homonyms, and stemming. Therefore, system using metadata for represent an article content, and the search results can more precise [4].

This study using Diabetes Mellitus Classification based diabetes disease diagnosis from World Health Organization Geneva. Ontology of diabetes disease is constructed using the Protege OWL format, then be weighted by density. System using weighted tree similarity algorithm to measure similarity. This application search convergent information by defining the concept of input conditions intended by the patient. The result of this research is the application that can help to diagnose diabetes early and as a guide for people with diabetes in the monitoring of disease.

II. LITERATURE REVIEW

A. Ontology

In heterogeneous environment, it is very important to use domain ontology as a reference for ontology processing. Domain ontology will help to solve heterogeneity problem in the case of concept misperception [5].

Ontology is the fundamental form of representation knowledge about the real world. In a review of Computer Science, Ontology has several meanings:

1. Ontology is an explicit specification of conceptualism [6];
2. Ontology defines the relationships, concepts, and other differences that are appropriate for modeling the problem domain [7];
3. Specifications shaped the definition of representational vocabulary like classes and relationships, which gives meaning to the formal constraints and vocabulary on the use of coherent [7].

Ontology is a symbolic representation of the knowledge about the object, class of objects, object properties, and relations between objects to represent knowledge of the application domain.

B. Web Ontology Language (OWL) Format

Web ontology language (OWL) is a Web-based ontology language that was designed for the purposes of integration and interoperability-related documents on the Web. OWL can provide additional vocabulary with a formal semantics [8]. It's better than RDF, which does not have a description of the relationship. OWL can explain or describe the semantics of the property and the class of a document, as well as how the association. OWL currently has three categories of languages, namely OWL Lite, OWL DL, and OWL Full. As an ontology language, OWL is used for various purposes, ranging from manufacturing to the class definition in a computer program, making e-commerce applications, to making search tool.

C. Weighted Ontology

According to [9], the calculation of the weight of concept ontology for this study are based on the density factor. The more a concept related to other concepts, the concept of the greater weight. Calculation of weight concepts in the ontology using the formula:

$$W(c) = \frac{in\ degree(c) + out\ degree(p)}{in\ degree(O) + out\ degree(O)} \quad (1)$$

W(c) = weight of concept c;
in-degree (c) = the total number of relation that goes into the concept c;
out-degree (p) = the total number of relation that out of the concept c;
in degree (O) = the total number of relation that led to the whole layered structure;
out-degree (O) = the total number of relation that leave the hierarchical structure;

D. Weighted Tree Similarity Algorithm

Weighted tree similarity algorithm is used to measure similarity of two pieces of tree/ontology [10]. Weighted similarity tree was originally a dealer on e-Business to match or get supplies sellers and buyers want most similar through weighted tree as a form of representation. In the marketplace, buyers and sellers advertise their needs and offer products / services. To get a match between the two, calculation of similarity between supply and demand should generate a list of sequential resemblance to buyers and sellers.

Weighted tree has a concept node labeled, arc weighted, and arc labeled to represent the parent-child relationship of an attribute of the product / service. Arc shows attributes labeled products / services, and the weight of the arc indicates the level of interest from the arc. Tree shapes previously only had labeled nodes. Calculation of weight concepts in the ontology using the formula:

$$\text{Similarity}(X, Y) = \sum A(Si)(Wi + W'i)/2 \quad (2)$$

Determination of the value of similarity is generally done by determining the level of similarity. Two weighted tree similarity ranged from 0 to 1. Similarity will be 0 if the two have nothing in common at all and if the same value is 1 [11].

III. RESEARCH FRAMEWORK

A. Construction of Knowledge

The construction of knowledge used in this research is shown in Fig.1.

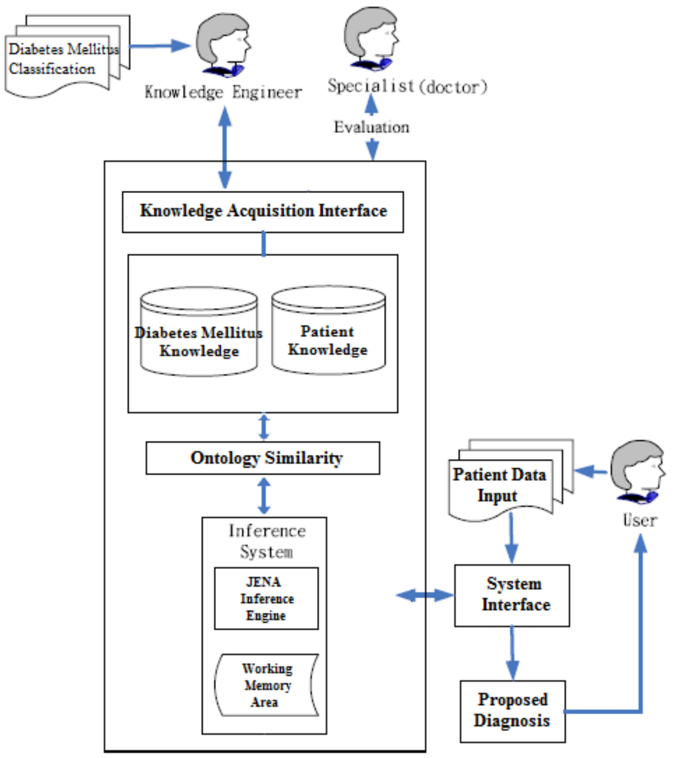

Fig.1. Construction of knowledge

This system (Java System Interface) receives input patient data from user. Then, system will build the patient ontology to represent patient knowledge. We are connecting Java applications to Protégé using OWL API. Then, system will calculate the weight of an ontology based on density. This system use JENA Inference Engine and working memory area for reasoning. The system would then do process similarity matching with Ontology Diabetes Mellitus using weighted tree similarity algorithm. Ontology has the highest similarity value will be the proposed diagnosis.

B. Patient Data Input

In this system, user enter patient data such as name, gender, age, results of tests (glucose, insulin, height, weight, diet, exercise, and psychological conditions).

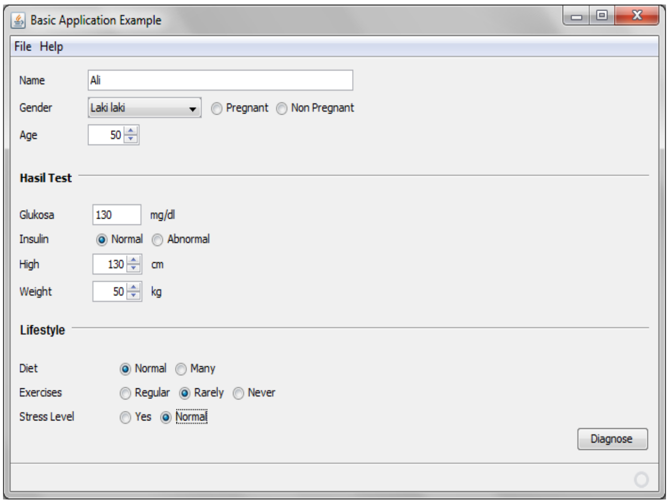

Fig. 2. Patient data input form

In this study, system classifies variables patient's data before process build ontology. Classification process is necessary to ease the process of ontology similarity. System classifies variables patient's data as follows: age (table I), glucose levels (table II), and body mass index (table III).

TABLE I
CLASSIFICATION OF PATIENT AGE

Classification	Age (years)
Children	0-12
Teenager	13-30
Adult	> 30

TABLE II
CLASSIFICATION OF GLUCOSE LEVEL

Classification	Mg/dL
Low	< 70
Normal	70 < x < 120
High	> 120

TABLE III
CLASSIFICATION OF BODY MASS INDEX

Classification	Index
Underweight	<= 18,5
Normal weight	18,5 < x < 24,9
Overweight	>= 25

Calculation of Body Mass Index (BMI) in this research using the formula:

$$BMI = \frac{\text{mass (kg)}}{(height(m))^2} \tag{3}$$

C. Build Ontology

Our research used Protégé 4.2 to build patient ontology and Diabetes Mellitus ontology (illustrated in Fig.3). Protégé produces ontology in OWL or RDF format [12]. We used OWL API to connecting Protégé to Java applications, and to find exhaustive information from knowledge represented in ontology [13]. Diabetes Mellitus Classification from World Health Organization Geneva used to represent Diabetes Mellitus knowledge. Patient data used to represent patient knowledge.

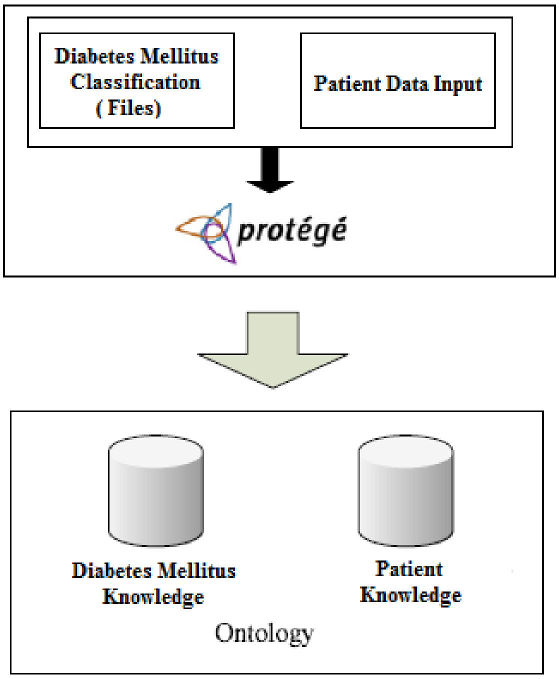

Fig. 3. Processes build ontology

This study using Diabetes Mellitus Classification based diabetes disease diagnosis from World Health Organization Geneva [14] are 4 kinds as follows:

1. Diabetes Mellitus Type 1 (Fig.4)
2. Diabetes Mellitus Type 2 (Fig.5)
3. Diabetes Mellitus Type Gestational (Fig.6)
4. Non Diabetes Mellitus (Fig.7)

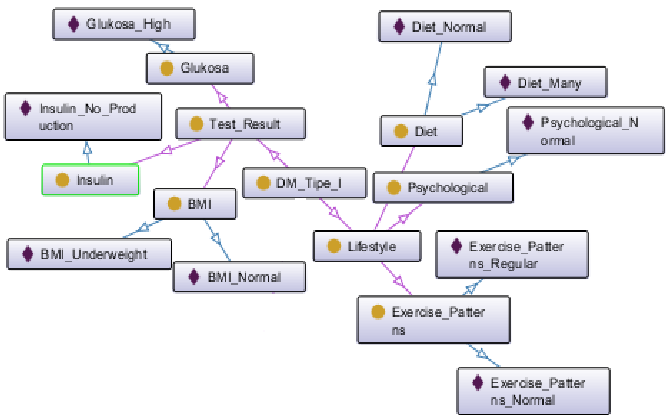

Fig. 4. Ontology Diabetes Mellitus Type 1

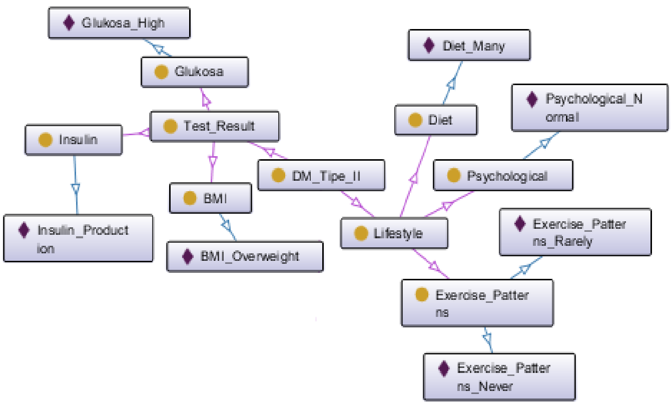

Fig. 5. Ontology Diabetes Mellitus Type 2

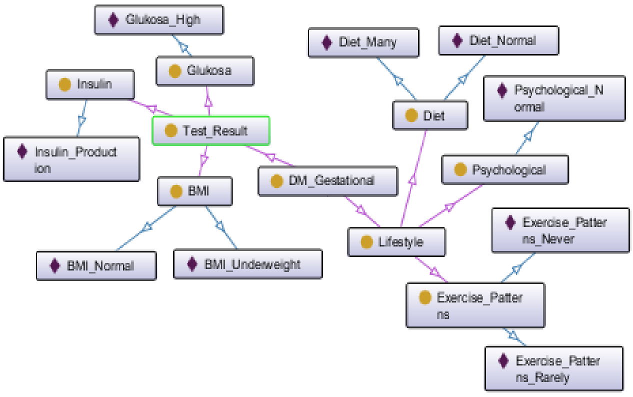

Fig. 6. Ontology Diabetes Mellitus Gestational

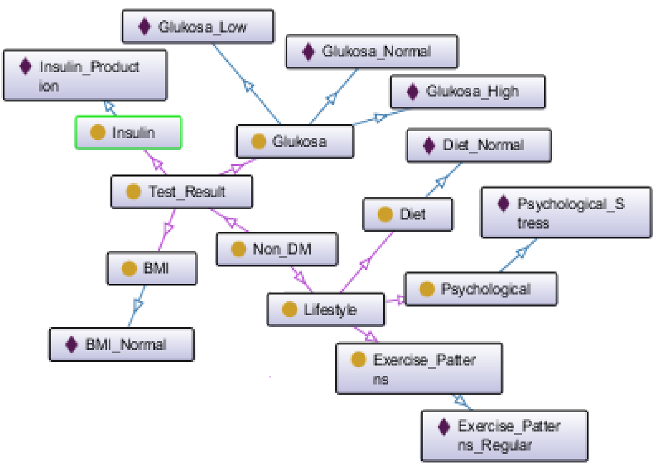

Fig. 7. Ontology non Diabetes Mellitus

D. Weighted Ontology Based on Density

Once the data into quantitative parameters, the system will represent these parameters into ontology (Fig.8) with the same structure as the master ontology, and then calculate the weight of each entity based on density. Calculation of weight concepts in the ontology is using formula (1).

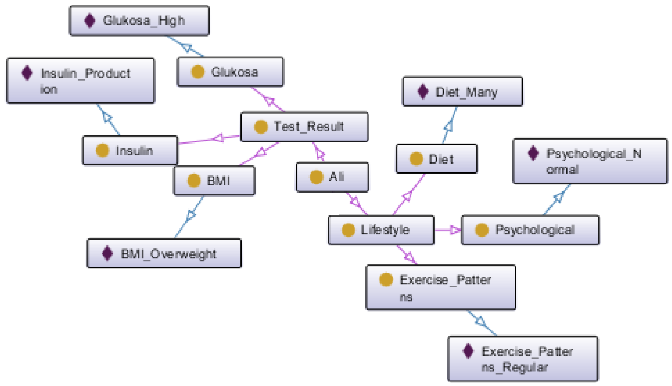

Fig. 8. Ontology patient Ali

W (Test Result) = (1 + 3) / (2 + 6) = 4/8 = 1/2
W (Lifestyle) = (1 + 3) / (2 + 6) = 4/8 = 1/2
W (Glucose) = (1 + 1) / (3 + 3) = 2/6 = 1/3
W (Insulin) = (1 + 1) / (3 + 3) = 2/6 = 1/3
W (BMI) = (1 + 1) / (3 + 3) = 2/6 = 1/3
W (Diet) = (1 + 1) / (3 + 3) = 2/6 = 1/3
W (Psychological) = (1 + 1) / (3 + 3) = 2/6 = 1/3
W (Exercise Patterns) = (1 + 1) / (3 + 3) = 2/6 = 1/3

E. Ontology Similarity

Once the ontology has the weight of each entity, the system will calculate similarity of this patient data ontology with Diabetes Mellitus ontology, i.e. Diabetes Mellitus Type 1 (illustrated in Fig.9), Diabetes Mellitus Type 2, Diabetes Mellitus Gestational, and non Diabetes Mellitus. Formulation of Weighted Tree Similarity calculation algorithm is using formula (2).

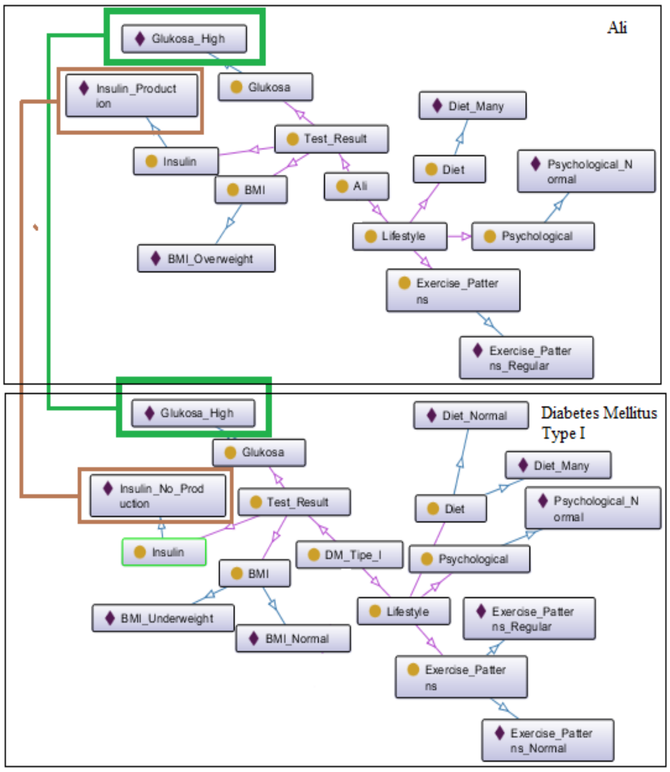

Fig. 9. Ontology Similarity (Ali, DM Type 1)

Similarity(Ali, DM Type1) = Sim("Glucose High", "Glucose High")*[(weight"Glucose"t1)+(weight"Glucose"t2)/2]+ Sim("Insulin Production, "Insulin No Production")*[(weight"Insulin"t1) + (weight"Insulin"t2)/2] + Sim("BMI Overweight, "BMI Underweight") * [(weight"BMI"t1) +(weight"BMI"t2)/2] + Sim("Diet Many, "Diet Many")*[(weight"Diet"t1)+(weight"Diet"t2)/2]+Sim ("Psychological Normal", "Psychological Normal") *[(weight"Psychological"t1)+(weight"Psychological"t2)/2]+ Sim("Exercise Pattern Regular", "Exercise Pattern Regular")*[(weight "Exercise Pattern"t1) + (weight" Exercise_Pattern"t2)/2]

Similarity(Ali, DM Type 1) = 1*[(1/3+1/3)/2] + 0*[(1/3+1/3)/2]+0*[(1/3+1/3)/2]+1*[(1/3+1/3)/2]+]+1*[(1/3+ 1/3)/2]]+1*[(1/3+1/3)/2]

Similarity (Ali, DM Type 1) = 4/6 = 0.67
Using the same calculation, the results of the other ontology similarity is following:
Similarity (Ali, DM Type 2) = 0.83
Similarity (Ali, DM Gestational) = 0.67
Similarity (Ali, Non DM) = 0.5

TABLE IV
ONTOLOGY SIMILARITY RESULT

Similarity	DM Type 1	DM Type 2	DM Gest	Non DM
Ali	0.67	0.83	0.67	0.5

From ontology similarity result (table IV), a value of ontology which has the highest similarity value is the reference determination of the type of diabetes disease. For example the case of Ali's patients (Fig.10), the highest similarity is Diabetes Mellitus Type 2 with similarity value of 0.83.

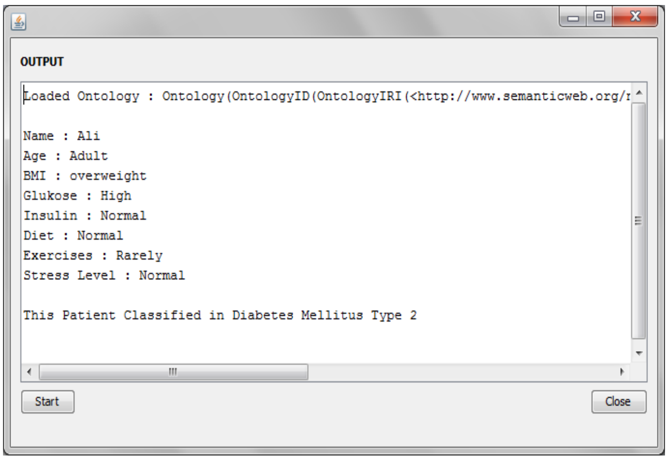

Fig. 10. Output System

F. Evaluation System

We let a specialist (doctor) check the consistency of our system. Disease diagnosis data that has been collected will be tested with the kappa index testing. Kappa index test is used to test the consistency of the result our system. Kappa index values are based on expert assessment that assesses the results of the test data according to their knowledge. Kappa index was calculated using the equation:

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (4)$$

P (A) is the proportion of how many times the experts agree and P (E) is the proportions of experts only by chance agree.

The system performed 100 patient data tests to check consistency of our system. The proportion of how many times the doctor agrees is 0.9355. The proportions of doctor only by chance agree 0.065. The value of Kappa index is 0.931. The consistency of our system based on doctor assessment that assesses the results of the test data according to their knowledge is 93.1%.

IV. CONCLUSION

From the evaluation performed at diagnosis system, the consistency of our system is 93.1%. Results of this study show that the representation in the form of OWL ontology using weighted ontology and weighted tree similarity algorithm can be used to represent knowledge about diabetes mellitus. This system can help to diagnose diabetes early and as a guide for people with diabetes to monitor the disease. Ontology supports the search of data / information by defining the concept of convergent intended by the user. Further research can be done by strengthening diabetes mellitus ontology in detail to improve the consistency of system.

ACKNOWLEDGMENT

This work is supported by Program of Informatics and Computer Science Universitas Brawijaya Malang, Department of Informatics Institut Teknologi Adhi Tama Surabaya, and the Doctoral Program in Department of Informatics Institut Teknologi Sepuluh Nopember so that this work can be done in April 2013.

REFERENCES

[1] Rahmawati, U.; Suryani, E.; Mukhlason, A.;, "Pengembangan Repository Pengetahuan Berbasis Ontologi (Ontology-Driven Knowledge Repository) Untuk Tanaman Obat Indonesia," *JURNAL TEKNIK POMITS*, vol. 1, no. 1, pp. 1-6, 2012.

[2] Amborowati, A.;, "Model Ontologi Untuk Informasi Jadwal Kereta Api Menggunakan Protégé," in *Seminar Nasional Aplikasi Teknologi Informasi 2007 (SNATI 2007)*, 2007.

[3] Zhang, G.; Jia, S.; Wang, Q.;, "Construct ontology-based enterprise information metadata framework," *Journal of Software*, 5(3), pp. 312-319, 2010.

[4] Sarno, R.; Rahutomo, F.;, "Penerapan Algoritma Weighted Tree Similarity Untuk Pencarian Semantik," *JUTI*, vol. 7, no. 1, pp. 35-42, January 2008.

[5] Kunaefi, A.; Sarno, R.;, "Ontology Mapping For ERP Business Process Variations," *STMIK AMIKOM*, 2013.

[6] Gruber, Thomas R.;, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *Int. Journal of Human Computer Studies*, vol. 43, pp. 907-928, 1995.

[7] Gruber, Thomas R.. (2007) What is an Ontology? [Online]. http://www.ksl.stanford.edu/kst/what-is-ontology

[8] Yang, D.; Miao, R.; Wu, H.; Zhou, Y.;, "Product configuration knowledge modeling using ontology web language," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4399–4411, 2009.

[9] Sarno, R.; Ghozali, K.; Nugroho, B.; Hijriani, A.;, "Semantic Matchmaking using Weighted Directed Acyclic Graph," *Institut Teknologi Sepuluh Nopember (ITS)*, pp. 1-6, 2011.

[10] Anistyasari, Y.; Sarno, R.;, "Weigthed Ontology for Subject Search in Learning Content Management System," in *IEEE*, Bandung, 2011, pp. 1 - 4.

[11] Sarno, R.; Yang, L.; C., Bhavsar;, "The Agent Matcher architecture applied to power grid transactions". Halifax, Canada: University of New Brunswik, 2003.

[12] Dewandono, R.; Fauzan, R.; Sarno, R.; Sidiq, M.;, "Ontology and Process Mining for Diabetic Medical Treatment Sequencing," in *International Conference on Information & Communication Technology and Systems*, Bali, 2013.

[13] Pramono, D.; Setiawan, N.; Sarno, R.; Sidiq, M.;, "Physical Activity Recommendation for Diabetic Patients Based on Ontology," in *International Conference on Information & Communication Technology and Systems*, Bali, 2013.

[14] World Health Organization, "Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications". Geneva: Department of Noncommunicable Disease Surveillance.