Ontology and Semantic Matching for Diabetic Food Recommendations

Achmad Arwan
Department of Informatics
University of Brawijaya
Malang, East Java, Indonesia
arwan@ub.ac.id

Bayu Priyambadha
Department of Informatics
University of Brawijaya
Malang, East Java, Indonesia
bayu_priyambadha@ub.ac.id

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

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

Heri Kristianto
Department of Nurses & Medical Surgery
University of Brawijaya
heri.kristianto@ub.ac.id

***Abstract*—Foods recommendation for diabetes patients is indispensable for controlling blood sugar levels. Currently, the foods preparation is done by a nutrition expert. The patient's dependence on the nutrition experts is very high, thus the selection of foods could not be done independently. The Automation system to determine foods combination for diabetic patients is needed to solve these problems. In this study, the automation system has been designed and implemented. The technologies used in this research are the OWL and SWRL. There are few researches that explore an automation process of foods recommendation for diabetes patients using the technology of OWL and SWRL. Domain knowledge based on Ontology is needed to process foods composition automatically. However, using SWRL and OWL technology is not enough, because the accuracy of the words required. A semantic ontology understanding was added using weighted tree similarity method to specify the composition of foods for diabetic patients. 73% data were able to be correctly predicted by this method.**

***Keywords*— Diabetes Mellitus, Diet, Ontology, Foods Recomendation**

I. INTRODUCTION

One of type of diabetes that has quick development process is type 2 diabetes [1]. Type 2 diabetes or Non-Insulin Dependent Diabetes Mellitus (NIDDM) can strike anyone regardless of age, generally the more attacking those who have overweight or obese. The early symptoms of type 2 diabetes are difficult to detect. This is indicated by one of three people who have type 2 diabetes do not know they have the disease. Type 2 diabetes is a chronic disease that occurs because the pancreas is unable to produce enough insulin (the hormone that regulates blood sugar) or when the body cannot effectively use the insulin that is produced. This condition can lead to high blood sugar levels and can cause serious damage to many body systems, especially the nerves and blood vessels. Diabetes can lead to serious complications such as heart disease (cardiovascular disease), neurological disease (diabetic neuropathy), kidney failure (diabetic nephropathy), and eye disease (diabetic retinopathy). For people whom suffer from type 2 diabetes controlling the food consumed is the important thing to do. Because, by controlling their food, patient of type 2 diabetes can control the condition of sugar level in their body.

Dietary control in diabetes patients is a very important. Nutritional foods is very diverse. Lack of information on nutritional diet on diabetic patients can make the condition of patients get worse. Diet in diabetes patients can be used as an attempt to control sugar levels in the blood.

Today, patient of diabetes consult about foods nutrition only with expert nutrition. Patients often experience of rising sugar levels suddenly only because of not knowing the levels of nutrients that ate. For non-medical patient, nutrition level is an information that they do not see in daily life. With this condition, there are needed some actions to make all patient have a knowledge about food and nutrition. To solve this problem, it is required a tool that make a patients can decide their menu by them self, and it will decrease the dependencies of patient of diabetes on the nutrition expert.

An automatic approach is one method that will propose in this research. This study proposed the construction of ontology-based foods recommendations for type 2 diabetes patients using ontology and semantic matching. Several researchers have conducted research on ontology to set food menu. Ontology is built as a domain knowledge about food menu which previous presented on [2]. The mechanism of that system is similar to knowledge-based system. There are knowledge and condition that will be queried into the knowledge. And then system will show the diet menu for patients. In this research, proposed some enhancement at the existed approach to increase the sense of semantic in the system. Some semantic search method will be adopted.

978-1-4799-0425-9/13/$31.00 ©2013 IEEE

II. RELATED WORK

A. Previous Method

Several researchers have conducted research related to the determination of foods menu for diabetic patients. The purpose of some previous researches are to find a method that automatically determine the foods menu.Nutritional therapy is essential to avoid, control and slow down the process of complications in diabetic patients [3]. In such research, has been declared a method in the preparation of the ontology automatically with hierarchical clustering approach.

Automation determination system has been built using ontology and combination ontology with fuzzy algorithm known as fuzzy ontology [4]. The focus of those research was in diabetes cases [5][6][7] and the rest is in document searching [8]. There are uncertain a data value not precision data on a domain of knowledge. Fuzzy algorithm was used to solve the uncertainty.

Control system construction of foods for diabetic patients was conducted in [9]. In this study, there is a process of reasoning within ontologies. The ontologies which used in the study are foods ontology and menu ontology.

B. Analysis of Previous Method

Several researchers were conducted in several approach of constructing menu for diabetes patients. Some automation methods have been studied using ontology to represent the knowledge. There were some data that has been used for testing phase. Data were taken from several real diabetes patients. Then, the data was used as an object query to the ontology of domain knowledge. The result is shown as the arrangement of menu for diabetes patient.

The problems were found in previous approaches. System that uses the ontology as domain knowledge will be functioned if data is exact and contained in ontology. If the data is not contained by ontology, the system will not show the result. In the other side, there are possibilities that the data is not available in the ontology, but the system must be show out the result. The biggest question is how to solve this problem. There is a method called Weighted Tree Similarity that used to count degree of similarity two tree data [10]. In [11], the method is used to count degree of similarity of agents in e-business environment. Agents are represented as node-labeled tree, arc-labeled tree, or arc-weighted tree.

This research will propose a method to construct diabetes menu using ontology and semantic matching among ontologies using weighted tree similarity. The method and the direction of solving will be explained further.

III. RELATED THEORY

A. Daily Calorie Needs

To determine the nutritional recommendations on diabetes mellitus (DM) patient required daily calorie needs calculated in accordance with condition of the patients and the type of their activities [2]. Daily calorie needs is calculate by finding the desired weight using the Body Mass Index (BMI).

$$BMI = \frac{mass(kg)}{height(m)^2} \tag{1}$$

Then the calculation of BMI is used to classify whether patients fall into skinny, normal, or fat category.

The activities of the patient are also a supporting factor in determining daily calorie needs [2]. There are 3 types of daily activities of patients which are light, medium& heavy. The following types of activities the patient based on the work is shown on Table 1.

TABELI.
THE TYPE OF PATIENT ACTIVITIES

Light	Medium	Heavy
Clerk's Office	Student	Sailors
Store clerk	Light industrial employees	Labour
Teacher	Housewife	Dancers

Next is the metric table calories needs per kg according to the BMI calculation type of patient activities [2].

TABELII.
MATRIX NEEDS CALORIES PER KG IDEAL BODY WEIGHT

	Activity		
BMI	***Light***	***Medium***	***Heavy***
Obesity	25	30	35
Normal	30	35	40
Underweight	35	40	40-50

The calorie intake per day can be described in the following formula.

$$Calory = mass(kg) * matrix \tag{2}$$

To find out the ideal weight, the following formula is used.

$$W = (heigth - 100) - 10\% \tag{3}$$

For example patient with a weight of 100 kg and a height of 170 cm with the job as a teacher. The body mass index can be counted using the aforementioned formula as follow 100 / $(1.7)^2$ = 34. In the matrix of BMI, 34 are classified as obesity (over weight). To count ideal weight, we can count using formula 3, so the ideal weight can obtain as 63 kg. Using formula 3, there are some conditions, for patient that has gender as a man and has height greater than 160 has to decrease by 10% of height. If the patient has height less than or equal to 160 the calculation is be done without using decrease by 10% of height. On the other hand, female patients and has height greater than 150, it has to decrease by 10% of height, and otherwise if less than or equal to 150.

To count the ideal amount of calories consumed in a day, the number of calories per kg body weight multiplied by ideal body weight [2]. To find out the number of calories per kilogram of body weight, we can refer to the matrix needs calories per kg of body weight by combining BMI status and activity level of the patient. So we can obtain 25 calories, because the example is a teacher and has the nutritional status of obesity. Occupation as teacher falls into the category of light activity. And to calculate calorie needs for a day, then the number of calories per kg ideal body weight multiplied by ideal weight, therefore the example patient is 1575 calories.

B. Ontology

The ontology is used to describe a concept that has the characteristic property & attribute [12]. Ontology consists of elements, attributes, class and object [13]. OWL is Web Ontology Language or a language for defining web ontology. Meanwhile, ontology is the science that is used to describe a variety of entities that exist in the world and how to represent them related [11][12]. OWL is built based on RDF schema. OWL and RDF has the same basic function: defines the classes, properties, and relationships. Nevertheless more than capabilities offer OWL RDF in defining complex relationships such as subclass, restriction, disjoint, constraint, and cardinality.

IV. METHODOLOGY

The methods used in this study are develops of foods ontology, developing of foods calorie ontology, developing rule and search foods menu. All the steps are figured in Fig. 1.

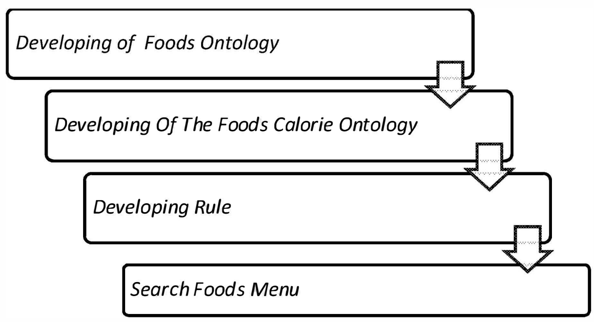

Fig.1.Research framework

A. Develops of Foods Ontology

Develops foods ontology is the most basic step in this research. Foods ontology is a an overview of knowledge about foods that includes of concepts, relations, attributes and individual foods. Knowledge of foods is reflected in the foods ontology which is a set of core knowledge in the search menu list for diabetes patient.

Knowledge of foods taken from some reference of foods and enriched with the information from nutrition expert [2]. All information about food reserved only for diabetic patients. The foods ontology is represented in Fig. 2.

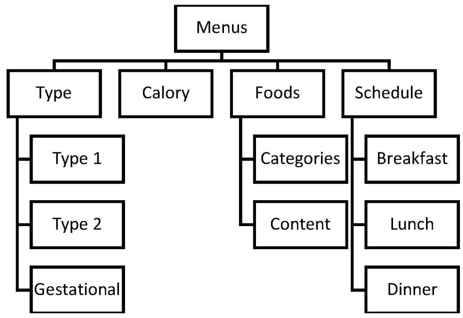

Fig.2.Foods Concept Hierarchy

B. Develops Of The Foods Calorie Ontology

Calories foods ontology is a knowledge representation that consists of information of diabetic patients and the number of calories of foods a day. Calorie foods ontology is spread from 1300 calories to 2300 calories. Hierarchical structure of the foods calorie ontology is depicted in Fig.3. This ontology will be used to process the matching of ontology of patients which is entered as the input system. The matching process occurs at the stage of search foods menu.

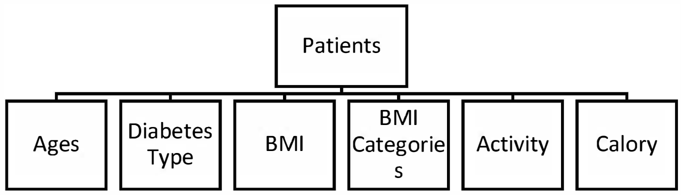

Fig. 3.Patients calorie needs ontology hierarchy

C. Develops the Rule

Rule in the ontology of serves as an inference engine. Rule served to do a classification or grouping the data categories in the ontology. In this study the language used as the definition of SWRL rule ontology.

SWRL is a combination of OWL-DL with RuleML [6] which was modelled in a Horn-claouse. Horn-clause is a representation of a conditional if-then that is more familiar. SWRL rule can be either C (x), P (x, y) SameAs (x, y) where C is an OWL description, P is an OWL property and x, y are the individuals or the data on the OWL. Examples of SWRL rules are as follows:

hasParent (?x1,?x2) ∧ hasBrother (?x2,?x3) ⇒ hasUncle (?x1,?x3)

Based on SWRL above can be interpreted as the following, x1 hasParent x2 and x2 hasBrother x3 then x1 hasUncle x3.

In this research the rule was made to classify the data automatically. Datas are grouped based on the food content. Some of the rules used are described as follows:

Food(?f), contain(?f, Vitamins_and_Minerals), hasKind(?f, "Fruit") -> Fruits(?f)

Food(?f), contain(?f, Vegetable_Protein) -> Vegetable_Protein_Foods(?f)

Food(?f), contain(?f, ?n) -> contained(?n, ?f)

Food(?f), hasKind(?f, "Milk") -> Milk(?f)

Food(?f), contain(?f, Animal_Protein) -> Animal_Protein_Foods(?f)

Food(?f), contain(?f, Fiber), contain(?f, Vitamins_and_Minerals) -> Vegetable_Foods(?f)

Food(?f), contain(?f, Carbohydrate) -> Carbohydrate_foods(?f)

D. Search Foods Menu

Search foods menu is done in several stages. The stages are receiving input ontology of patients, matching patient ontology with calorie foods ontology, and the final stage in the process of using SPARQL queries on foods ontology. The search flows of process illustrated in Fig.5.

To match the ontology we use Weighted Tree Similarity method [10]. Matching is performed on every leaf of the tree by considering the weight of each arc. Fig.4 shows a sample tree calculated similarity.

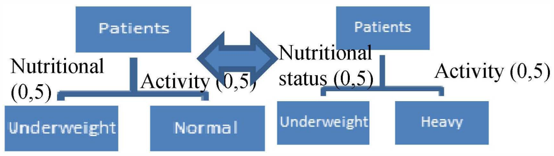

Fig.4. Example of Similarity Tree

Calculation of similarity of the tree above are:

Sim (nutrition) = 1 (0.5 + 0.5)/2 = 0.5
Sim (activity) = 0 (0.5 + 0.5)/2 = 0
Total = 0.5 + 0 = 0.5

Results calculation of the similarity of both the trees is 0.5. This method is used in the process of matching of ontology which is compare the patient's data and calorie. Both data is represented in tree. The highest degree of similarity resulted by this process will be chosen.

The next process is query the matching result to food ontology using SPARQL. The result will be showed the composition of food for diabetes patient. Fig. 5 will be described the flow of matching and searching of menu.

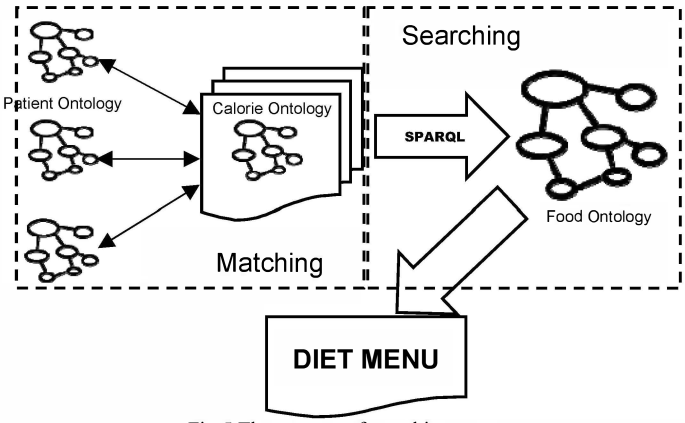

Fig.5.The process of searching menu

V. System Architecture

The proposed method will be implemented using several technology and tools. The input system or end user interface is implemented using Java Swing technology. In this layer, system is displayed as a desktop-based application. The second layer is the controller layer that act as a processor of the system. Several libraries are used in this layer. Jena and OWL API and JENA are used in this processor layer, beside java object oriented was conducted here. All of the tree or ontology in matching process is represented as an object. The last layer is the layer of data. This means that this layer act as storage of data. Data is described as an ontology that represents the domain knowledge. Ontology is built using tool named Protégé. Ontology contain of inference rule named SWRL. All of rules that define as SWRL was run or interpreted using inference engine. In this study, we employ the Pellet for reasoning engine. Pellet was equipped in Protégé, so we can easily use that feature. All of this description is represented in Fig. 6.

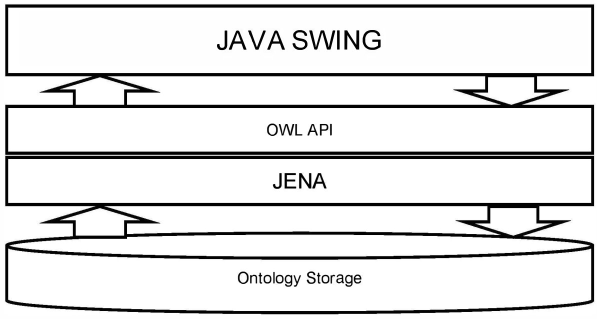

Fig. 6. System architecture

VI. Experiment

In the experiment of the search foods menu, the data used is the patient data including age, diabetes type, height, weight, nutritional status, activity and calories. Where in the process are the initial calculations prior to matching on ontology. Initial calculation does is to calculate the calorie and nutritional status, resulting in a system that was built, the values that should be included are age, diabetes type, height, and weight. The main display system is shown in Fig. 7.

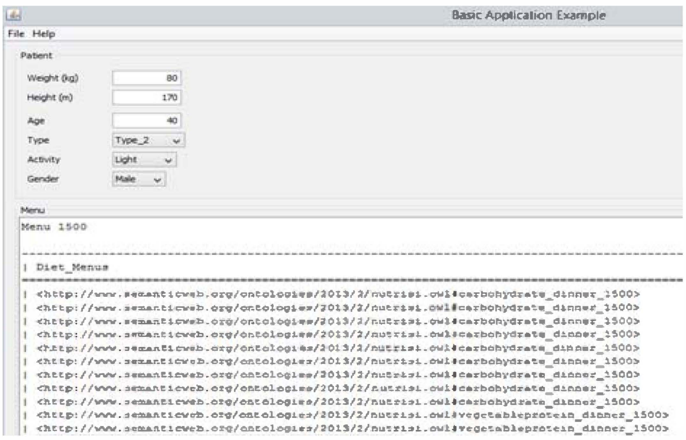

Fig.7. Main display system

The next process is the ontology matching of patients and calories foods. Ontology matching process is illustrated in Fig.8.

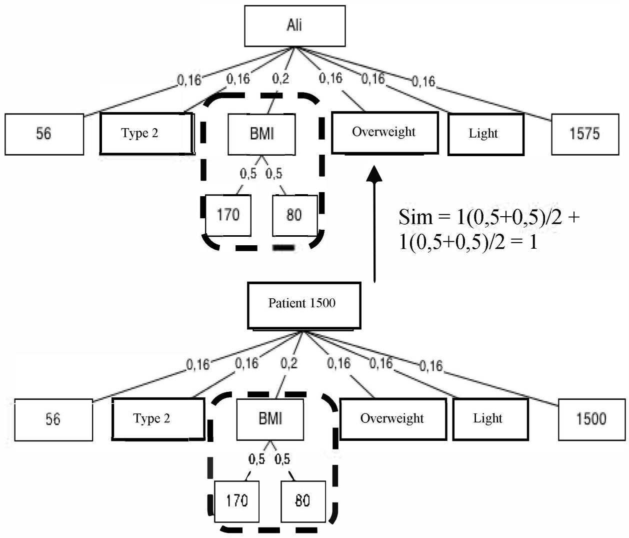

Fig.8. The first step ofsimilarity matching for menu

Fig.8 shows the description of the calculation process, where matching is done on the bottom branches. Explanation of the image above the calculation of similarity from the lowest branch is as follows:

$$Sim = 0.5\ (0.5 + 0.5) / 2 + 0.5\ (0.5 + 0.5) / 2 = 0.5$$

From the above calculation, obtained in common from the lowest node is 0.5. A value of 0.5 will put a value on it.

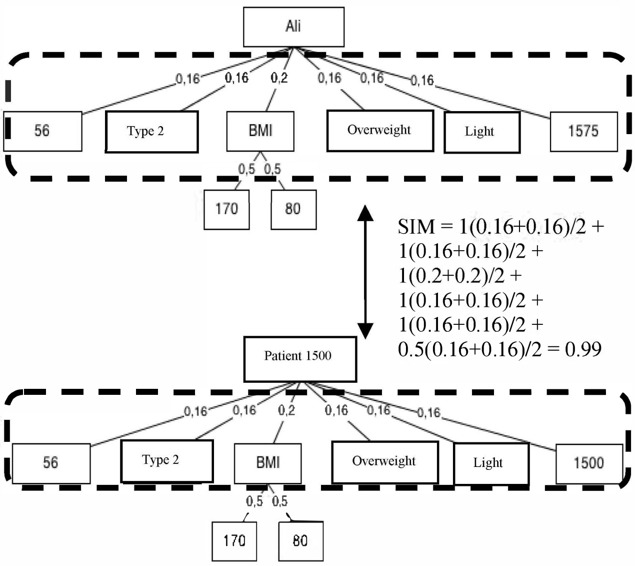

Fig.9. Second step of similarity matching for menu 1500

Fig.9 shows the continued process of calculation of matches from the points in the top level. From calculation the same way on the level below, then get that value matches from the patient's ontology and 1500 calories ontology was 0.99. There is likely value in common will be greater on the ontology of the other menus. Therefore the search process will be continued on menu ontology. And in the end found that, in the case of the example patient, a suitable menu for him is a menu with 1500 calories.

After the initial phase of ontology matching process is complete, then resume the search process the menu tree in the foods ontology. An overview of the foods ontology is reflected in Fig.10.

Fig.10. The foods ontology

The finding process of foods is done by running SPARQL in the ontology. The ontology has been equipped with reasoning that is defined using SWRL.

VII. RESULT

This section presents the results from the experiments. Ontology patient has 30 instances on diabetic patient ontology with vary conditions. Then through the user interface we have tested 30 patients with diabetes were randomized to look for individuals who have the highest level of similarity with it. Table III show the confusion matrix of this experiment.

TABELIII.
THE CONFUSION MATRIX OF EXPERIMENT

		Data Actual		Total
		True	***False***	
Prediction	***True***	(TP) 22	(TN) 0	(P') 22
	False	(FP) 8	(FN) 0	(N') 8
Total		(P) 30	(N) 0	30

Formula to calculate the level of accuracy of the data can be seen in the following formula.

$$ACC = \frac{TP+FN}{P+N} \quad (4)$$

Where, TP (True Positives) is the data that has true result and predicted as a true. TN (True negative) is data that is false but predicted as a true. FP (False positive) is the data is a true but predicted as false. And, FN (False negative) is data that is false and predicted as false. While P is the number of actual data of true, and N is the number of the actual false. So, it can be concluded that the accuracy of the defect detection system on this research is 0.733.

VIII. CONCLUSION

The determination of the list of menu foods for diabetic patients is crucial. List of foods menu based on the number of calories per day could help patients with diabetes to control their blood sugar levels. Setting the menu list can be done by building a foods ontology.

Search foods by using ontology can be done based on the context. Unlike the search in a relational database, if the data that is searched for is not found in the database, it will not generate an output. Semantic search method is used to calculate similaritas of data, so data doesn't have to be exactly same with the existing data. With the combination of some data then the context search will be formed on the method of semantic search.

Experiment results show that within 30 data sample of diabetic patients, system is able to recommend 22 data food recommendation (73%). It means that 73% of the data are correctly predicted. Further research will be focused on the methods of calculation of similaritas and accuracy improvement. Model calculations can be improved to make it more effective and efficient. Model ontology can also be improved, so that the process can be done quicker and more accurate.

REFERENCES

[1] G. Hu and J. Tuomilehto."Lifestyle and outcome among patients with type 2 diabetes" , Int. Congr. Series, vol. 1303. Aug. 2007. pp. 160–171.

[2] Sarwono .W,Gatut .S, Kartini .S,Rochamah .M, "*Cara mudah mengatur makanan sehari-hari seimbang & sesuai kebutuhan gizi*". FK University of Indonesia. 2004. ISBN 979-496-314-3.

[3] Li, H.C., Ko, W.M. "Automated Foods Ontology Construction Mechanism For Diabetes Diet Care". *IEEEMachine Learning and Cybernetics*. 2007.

[4] Lee, C.S., Wang, M.H., Hagras, H. "A Type-2 Fuzzy Ontology and Its Applicationto Personal Diabetic-Diet Recommendation". IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 18, NO. 2. 2010.pp. 374 – 395.

[5] Pramono, D., Setiawan, N.Y., Sarno, R. & Sidiq, M."Physical Activity Recommendation for Diabetic Patients Based On Ontology".*The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

[6] Dewandono, R.D., Fauzan, R., Sarno, R. & Sidiq, M. "Ontology and Process Mining For Diabetic Medical Treatment Sequencing".*The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

[7] Saadah, U., Sarno, R., and Yuhana, U.L. "Latent Semantic Analysis and Weighted Tree Similarity for Semantic Search in Digital Library". *The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

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

[9] Cantais, J., Dominguez, D., Gigante, V. "An Example of Foods Ontology for Diabetes Control". *Proceedings of the International Semantic Web Conference 2005 workshop on Ontology Patterns for the Semantic Web*. 2005.

[10] Sarno, R., Rahutomo, F. "Penerapan Algoritma Weighted Tree Similarity Untuk Pencarian Semantik". JUTI. Vol. 7, No. 1. 2008. Pp. 35-42.

[11] Sarno, R., Anistyasari, Y., and Fitri, R., "Semantic Search", Andi Publisher, 2012, ISBN 978-979-29-3110-5.

[12] Kuan M. "Using SWRL and OWL DL to Develop an Inference System for Course Scheduling". Master Thesis, Chung Yuan Christian University, Taiwan, R.O.C. 2004.

[13] Hebeler, J., Fisher, M., Blaze, R., Perez-Lopez, A. "Semantic Web Programming". Wiley Publishing Inc. 2009. Canada.