ONTOLOGY AND PROCESS MINING FOR DIABETIC MEDICAL TREATMENT SEQUENCING

Rahadian Dustrial Dewandono[1], Reza Fauzan[2], Riyanarto Sarno[3], Mohamad Sidiq[4]
[1234] Department of Informatics - Faculty of Information Technology
Institut Teknologi Sepuluh Nopember
Surabaya 60111
[1]dewandono11@mhs.if.its.ac.id, [2]fauzan.reza12@mhs.if.its.ac.id,
[3]riyanarto@if.its.ac.id, [4]sidiq12@mhs.if.its.ac.id

ABSTRACT

Diabetic treatments consider information related to both diabetes and several complications. To cure diabetes, it requires not only contextual data but also sequential data representing medical treatments that have been given.

This paper proposes a novel method for giving recommendation in cases of diabetic treatment. The method combines ontology and process mining. Ontology has an important role for matching contextual data of a patient with that of other healthy patients, while process mining is used for conforming medical record of the patient to process model of healthy patients.

The method yields three main benefits: (i) to retrieve a similar information about the patient; (ii) to get similar process model of healthy patients; and (iii) to give sequential diabetic treatments.

Keywords: Ontology, Semantic Query, Process Mining, Conformance Checking, Medical Treatment

1 INTRODUCTION

International Diabetes Federation (IDF) shows that more than 246 million people suffer for diabetes in the world and growing to 380 million by 2025 [1]. The issues of diabetes require big effort to solve. Several scientists have conducted researches devoted to diabetes medical treatments. As well as in Computer Science, numbers of researches have been conducted to provide solution for curing diabetes. Diabetes is such a complex treatment that it drives several computer scientists and engineers to build decision support system for diabetic medical treatments [2].

Context-Aware Information System empowers organizations to analyze contextual data based on ontology [3]. Previous research [4] proposes an ontology-based anti-diabetic drug information system. The system presents context-aware solution for anti-diabetic drug. It enhanced data retrieval regarding anti-diabetic drug. Nevertheless, the problem of diabetes requires many drugs for controlling blood. There are many combinations of anti-diabetic drug. Another ontology-based research in diabetes improves previous research by fixing the faults. The system increases knowledge by handling the complexity of such combinations [4].

Ontology applications are utilized for searching and decision support system in recent years. The Europe Union (EU) supports an ontology-based framework on constructing of domain knowledge for cardiovascular disease [5]. Ontology is a combination between Artificial Intelligence (AI) and machine language. It stores, shares, and reuses knowledge. It contains natural language processors and knowledge representation. Ontology is used for communication between human and computer system. Ontology is useful for information retrieval and knowledge management [6, 7, 8]. Utilizing an ontology empowers to find exhaustive information by matching the input data to the knowledge represented in ontology.

In cases of diabetic medical treatment, ontology is useful for representing knowledge about patient. There are some problems not able to be handled by Relational Database Management System (RDBMS). For instance, we want to find someone who has characteristics of diabetes type 1. In RDBMS, we should insert all of required data in order to get the result through exact matching. Nevertheless, sometime the data given as the input are imprecise. Indeed, it could not be tackled using exact matching in RDBMS. It might belong to either type 2 or type 1. In relational database, it is difficult to find this circumstance. In context-aware information system, such problem could be tackled. By using ontology, the system could find the person, even more than one person, since every characteristic has special relation, which does not exist in relational database. Therefore, ontology plays an important role for contextual data retrieval and semantic searching.

Nevertheless, diabetes does not only require context-aware solution. A recommendation of diabetic medical treatment is likely a sequence. The treatment of each person could be different based on several factors (e.g., the characteristic, recent condition, etc.). Furthermore, the relationship with other diseases has an important role for giving the treatment. Some diseases have contraindication with diabetes. It is highly hazardous if we give the treatment to person suffering such contraindication and diabetes at the same time. Therefore, it needs process-aware analysis for diabetes prognosis.

Process Mining is a technique to obtain knowledge from event log [9, 10, 11]. The event log is created from the system every time it executes certain activities. From the event log, we could obtain useful information (i.e., what health care activities were conducted by certain patients, in which department the patients were cured, when certain patients were cared, etc.).

In this research, we propose a novel method for diabetic medical treatment sequencing. The method combines context-aware technique toward patient ontology and process-aware analysis using process mining. In cases of medical treatment sequencing for diabetes patients, both ontology and process mining are used. Ontology is used as semantic query to acquire similar patients. Since sequence is an important point in diabetic medical treatment, such technique is not enough to support recommendation for diabetic patient. Sequential recommendation based on historical data should be provided. Therefore, we extend by proposing historical searching based on Process Mining. Process Mining could extract the knowledge in medical record of the patients. It is to compare the event log of the incoming patient to that of previously healthy patients. The objective is to find the most similar medical record of patient cured by the medical center. A context-aware event log referring to medical treatment ontology is used. It could support further investigation of medical treatment. Conformance checking results the fitness value between process models discovered from medical record of the incoming patient and the event log representing existing health care standard treatment based on historical data.

1.1 Related Diseases

Diabetes has complication. Several diseases might be prior or complication to diabetes. The diseases could exist due to diabetes. For instance, since a patient suffers for diabetes, the patient might have other diseases. Another case is that several diseases could exist before the patient suffers for diabetes and it becomes one of the factors causing diabetes. These circumstances show that sequential analysis is required in cases of diabetic treatment.

1.2 Process Mining

Process Mining is an emerging field located in the center of data mining and computational intelligence, as well as, process modeling and process analysis [10, 12, 13, 14]. The main objective is achieving knowledge from event logs. For further investigation, Process Mining has three domains: process discovery, conformance check, and enhancement. Process discovery perform building structure from event log. Conformance checking compares event log to corresponding process model. Enhancement is used as process analysis to improve performance of organization.

1.3 Conformance Checking

Conformance checking has a role for comparing an event log to corresponding process models. The objective is to measure fitness value representing the deviation of behavior [10]. Hence, the commonalities and discrepancies are investigated.

To measure fitness value, a Petri-Net model of process should be analyzed for each case. There are four tokens: consumed token, produced token, missing token, and remaining token. The fitness could be measured using formula as follows:

$$fitness(o, N) = \frac{1}{2}\left(1 - \frac{m}{c}\right) + \frac{1}{2}\left(1 - \frac{r}{p}\right)$$

The formula measures fitness value of case o in process N. In the formula, m denotes missing token, c denotes consumed token, r denotes remaining token, and p denotes produced token. The fitness value of an event log is the summation of the value for every case. The optimal value of fitness is 1. It means that the event log exactly represents the behavior of process model.

2 MODEL, ANALYSIS, DESIGN, AND IMPLEMENTATION

The main idea of the proposed method is how to combine semantic searching and conformance checking in cases of diabetic medical treatment sequencing. Figure 2 depicts methodology conducted in this research.

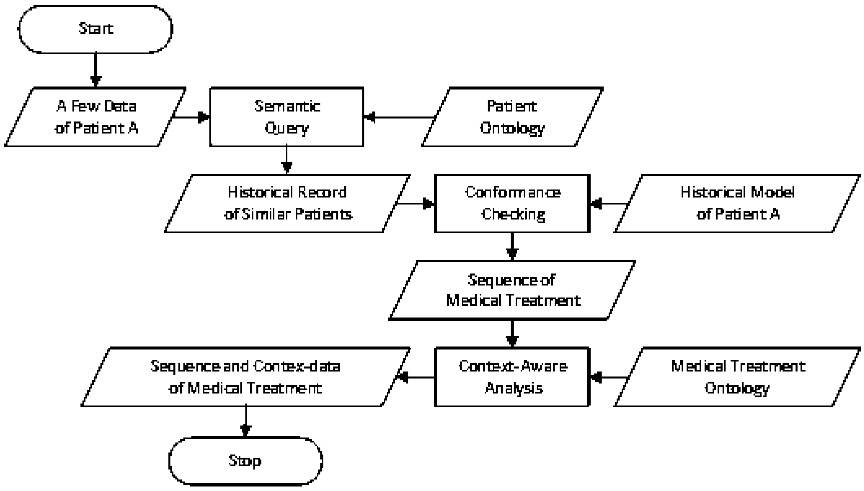


Figure 1. The Proposed Methodology in this Research

Few data about Patient A become input of the system. Semantic query is performed by involving patient ontology and such data. The objective is to get similar patients. Conformance checking is performed to gauge the fitness value between historical data of Patient A and historical data from these similar patients. The result is sequence of medical treatment. In order to add contextual information related to every activity of the sequence, context-aware analysis should be conducted. Finally, sequence of medical treatment with its contextual data is yielded.

2.1 Case Study

In this paper, we provide an example explaining how the proposed method works. Patient A gets regular care to Sehat hospital. Consequently, Sehat hospital has medical record about Patient A.

When Patient takes a regular care, it is revealed that he is diagnosed suffering for diabetes. Patient A wants to get recommendation for curing his illness. A clerk of hospital asks for information to Patient A. Patient A gives the clerk the data yielded from the regular care.

By utilizing patient ontology, the system measures the similarity between contextual data of Patient A to that of existing patients. By applying semantic searching [8], the system could retrieve contextual data from similar patients recorded in the system. The objective is to get contextual information related to the limited input given by Patient A.

For example, Patient A only has information about blood pressure, age, and weight. By comparing to related data, the system could represent Patient A as patient X, Y, or Z.

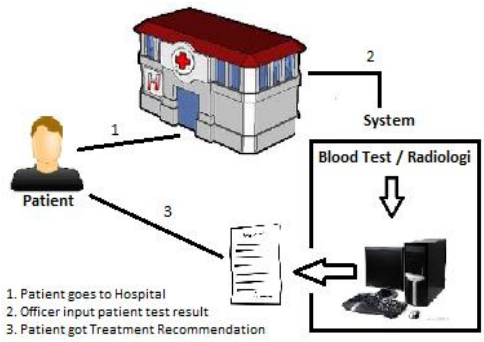


Figure 2. Illustration in cases of Patient A

To get more precise result, the system compares historical data of Patient A and process models built from historical data of patient X, Y, and Z. Patient X, Y, and Z were treated based on Standard Operational Procedure (SOP) as modelled in Figure. Nevertheless, each patient might have different process model based on the trace. The patient whose historical data is the best amongst the others in terms of fitness value is chosen as medical treatment reference. Finally, this step results a sequence of medical treatment for Patient A.

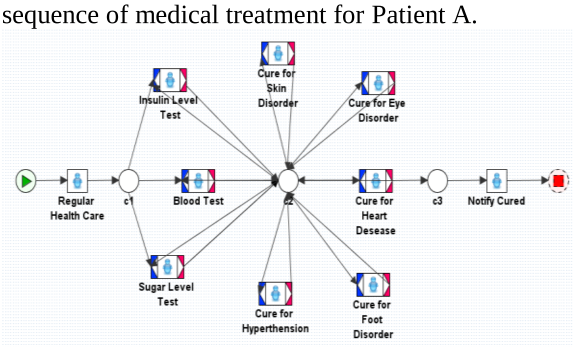


Figure 3. Process Model for Diabetic Medical Treatment

The sequence of medical treatment needs contextual information supporting every activity. For instance, the system could give the sequential recommendation as follows: Blood Test with type of test: Regular Blood Test; Cure for Skin Disorder with type: Vitiligo; and Notify for Cured with recommendation: Consume certain medicines.

Patient ontology consist of knowledge related to blood test information of Patient A given to the system. Patient ontology is constructed utilizing Protege 4.2. Protégé is on open software with Java-based platform. Protégé was developed from the Information Center at Stanford University and is used to access, create, and maintain the ontology.

2.2 Patient Ontology

The first step of the proposed methodology is designing patient ontology representing exhaustive knowledge about diabetic patient.

Figure 4 shows the preprocessing about this system divided into two sub-processing sections. Information extraction: The information includes patient test and characteristic, for instance blood pressure test, liver tests, renal insufficiency tests, diseases, age, gender, etc. Protégé builds knowledge: Protégé was used to construct the patient ontology in the preliminary experiment, and then the OWL DL format was adopted. In our study, ontology was constructed in the OWL DL (Description Logic). The classes of patient, attributes of patient class, and contents of patient properties were set up to construct the relevant knowledge of patient ontology.

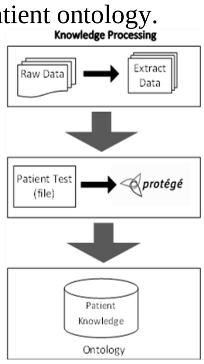


Figure 4. Knowledge Processing

Figure 5. Input Form

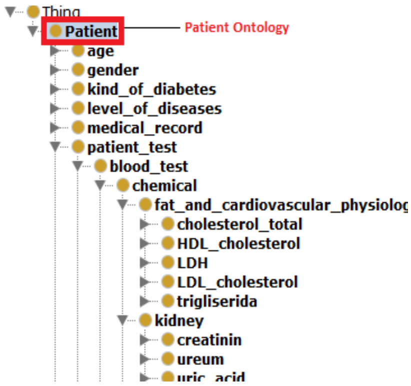


Figure 6. Patient Ontology

Figure 6 is the patient ontology created by Protégé. The frame represents patient knowledge and patient tests result. It shows the whole knowledge of patient ontology.

2.3 Weighting Patient Ontology

Blood test result is important for understanding other diseases. Figure 5 shows the input data. Patient A does not fill the entire field. Patient A is only required to fill several data that he/she has had.

We build an ontology from given information. However, the information could be imprecise. For instance, a normal result from cholesterol is 110 - 200 mg/dl. If the result is 215 g/dl or it is still within the threshold, it could be either normal or high for some reasons. We include fuzzification to solve it in the program we have. The result from fuzzification is used for weighting the new ontology from input data.

2.4 Ontology Matching

After the new ontology is completed with the weight from every result, it should be matched with patient ontology. The objective is to retrieve full information from patient ontology. Ontology matching is used to find other instances that have a close similarity with the new patient. It enables the system to analysis more about patient. Even though the input data is limited, the system could retrieve exhaustive data from existing patient ontology.

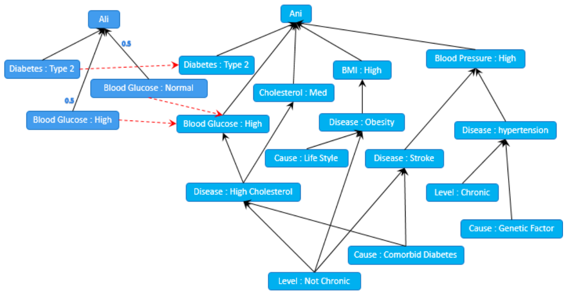


Figure 7. Ontology Matching

The difference of such structures is that the new ontology less level than patient ontology has. Therefore, before we calculate the similarity, we have to find the same parameter in level two using string matching. Furthermore, we calculate local similarity in level two using weighted tree similarity.

$$Similarity = \sum \left(\frac{A(s_i)(w_i + w_i^{'})}{2} \right)$$

We measure the similarity between the data of patients A to that of other healthy patients. The similar patients resulted from this comparison are chosen as the input in the next procedure. The similarity calculation result will be normalized to be 1-0. Patients are considered equal if they have threshold more than 0.8.

2.5 Conformance Checking

The role of process mining in this research lies in conforming medical record of Patient A with process models from some healthy patients. The healthy patients are chosen from ontology matching as explained in sub chapter 2.4.

Firstly, the system records every executed activity in the system. As depicted in Figure 8, the event log created contains Event ID, Case ID, Activity, Originator, and Time Stamp. To enable the system referring to ontology, activities are recorded as URL of other ontology (i.e., ontology for blood test, insulin injection, etc.).

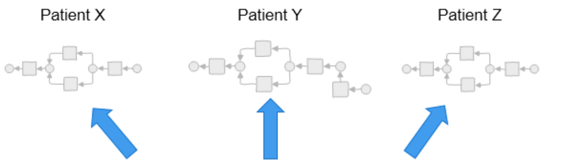
It is important to be noticed that the historical data discovered to process model are taken from the beginning of activity until the corresponding patients are diagnosed suffering for diabetes. The objective is to retrieve historical data, which is very close to historical data of Patient A.

Refer to Ontology for Medical Treatment

Event ID	Case ID	Activity	Originator	Time
1	1	RegularHealthCare.owl	Doctor C	11.25.33
2	1	BloodTest.owl	Doctor A	11.27.33
3	1	CureforSkinDisorder.owl	Doctor A	11.28.45
4	1	CureForHyperthension.owl	Doctor B	11.28.50
5	2	RegularHealthCare.owl	Doctor C	11.29.51
6	1	CureforFootDisorder.owl	Nurse	11.30.15
...	...	...	...	...

Figure 8. Medical Record of Patient A

From the event logs, process model of patient X, Y, and Z are discovered. Comparing them to historical record of Patient A would result various fitness values. The medical record, which has the highest value, is chosen as medical treatment sequencing. The sequences of treatment come from activities after being diagnosed suffering for diabetes. Figure depicts conformance checking between medical record of Patient A and historical process models of patient X, Y, and Z.



Medical Record for Patient A

Event ID	Case ID	Activity	Originator	Time
1	1	RegularHealthCare.owl	System	11.25.33
2	1	BloodTest.owl	Clerk	11.27.33
3	1	CureforSkinDisorder.owl	System	11.28.45
4	1	CureForHyperthension.owl	Clerk	11.28.50
5	2	RegularHealthCare.owl	System	11.29.51
6	1	CureforFootDisorder.owl	Nurse	11.30.15
...	...	...	...	...

Figure 9. Illustration of Process Conformance Checking

For instance, after conducting conformance checking, patient Z is chosen due to its medical record. In the medical record, it is recorded that patient Z performed "Cure for Skin Disorder" activity, "Cure for Eye Disorder" activity, "Regular Health Care" activity, and finally "Notify Cured" activity. Patient A is given such chain of activities until reaching "Notify Cured" activity.

2.6 Context-Aware Analysis for Further Investigation

The aforementioned methods only enable system to give sequences of medical treatments without any contextual data. These sequences are not enough to be presented as diabetic medical treatments. Therefore, we extend by analyzing each activity based on related ontology. The objective is to present exhaustive information about sequential recommendation.

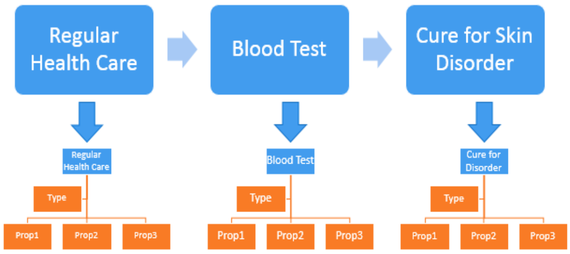


Figure 10. Related Ontology of Each Activity in Diabetic Medical Treatment Sequencing

We could analyze related information about certain activity in sequential recommendation by referring each activity to the related class in medical treatment ontology. Therefore, further recommendation (i.e., which type of blood test that the patient took, which part of regular test that the patient got, etc.). Such information is not stored in historical record of the patients.

3 RESULT

In this section, we present experimental design for the proposed method. Figure 13 describes whole scenario of evaluation. We use data set containing both sequential and contextual data related to diabetic medical treatment sequencing. The objective is to gauge the accuracy based on ROC framework. The scenario of evaluation in this paper is depicted in Figure 13

Firstly, after designing patient ontology, we compare Patient A to a complete set of patient ontology. In patient ontology, there are four patients having exhaustive information: Dewa, Rahardian, Reza, and Fauzan. Patient A is compared to those four patients in order to get similar patients. The comparison process is depicted in Figure 14.

Figure 12 (a) shows ontology made from input data from Patient A. In addition, we calculate the similarity from Patient A to the others.

Patient A : Dewa

$$Similarity = Cholesterol\ Normal + LDL\ Normal$$
$$Similarity = 1 \times [(0.25+1)/2] + 1 \times [(0.1+1)/2]$$
$$Similarity = 1.175$$

Patient A : Rahardian

$$Similarity = Cholesterol\ High + LDL\ High + Systole\ High$$
$$Similarity = 1 \times [(0.75+1)/2] + 1 \times [(0.9+1)/2] + 1 \times [(1+1)/2]$$
$$Similarity = 2.825$$

Patient A : Reza

$$Similarity = DiabetesType1 + Cholesterol\ Normal + LDL\ High + Systole\ High$$
$$Similarity = 1 \times [(3+3)/2] + 1 \times [(0.25+1)/2] + 1 \times [(0.9+1)/2] + 1 \times [(1+1)/2]$$
$$Similarity = 5.575$$

Patient A : Fauzan

$$Similarity = DiabetesType1 + Cholesterol\ High + LDL\ High + Systole\ High$$
$$Similarity = 1 \times [(3+3)/2] + 1 \times [(0.75+1)/2] + 1 \times [(0.9+1)/2] + 1 \times [(1+1)/2]$$
$$Similarity = 5.825$$

From the calculation, the system results Reza and Fauzan as similar patient compared to A. Therefore, they are chosen as the patients whose process model is conformed to medical record of Patient A.

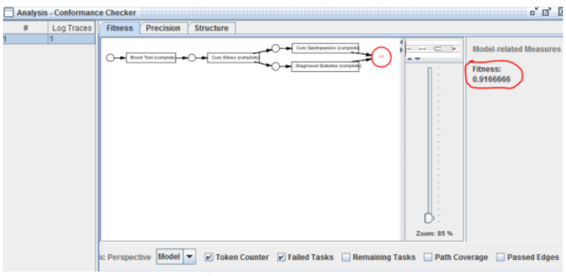


Figure 11. Conformance Checking in ProM 5.2

From the conformance checking, Fauzan has the most similar medical record to that of Patient A. Therefore, process model of Fauzan starting from “diagnosed diabetes” to “notified cured” becomes a sequential recommendation for Patient A.

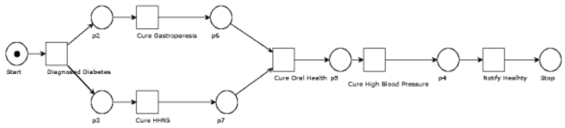


Figure 12. Process Model of Fauzan Given as the Diabetic Treatment for Patient A

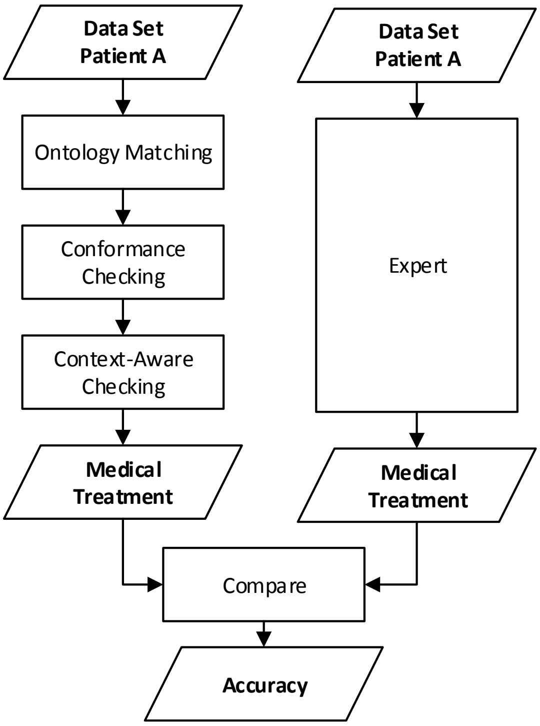


Figure 13. Scenario of Evaluation

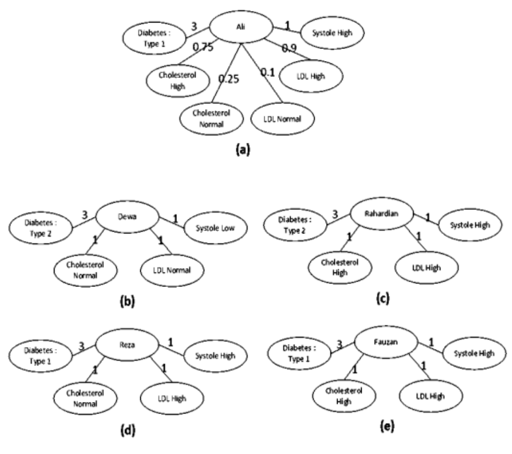


Figure 14. Ontology Matching Example

4 CONCLUSION

Diabetic is such a chronic disease that requires not only analysis about contextual data but also analysis about process models. Since diabetes might cause other diseases or is affected by other illnesses, analyzing medical records of the patient plays an important role in giving diabetic treatment.

In this paper, we propose a chain of methods to give contextual and sequential recommendation for patients who are diagnosed to suffer for diabetes. The proposed method combines ontology and process mining. Ontology, which is built regarding diabetes, benefits to match contextual data of the patient. A limited data yielded from some medical tests conducted by the patient is used as an input to compare to a more advanced model of ontology from other healthy patients.

Conformance checking is used to find similar process models of healthy patients. In order to enable in-depth analysis, ontology representing each activity of the process is linked with the log. The result shows that the proposed method could give medical treatments for diabetes patient based on historical data and contextual data.

5 REFERENCES

[1] International Diabetes Federation, "Diabetes Atlas," Brussels, 2006.

[2] J.-x. Chen, "Diabetes care decision support system," in *Industrial and Information Systems (IIS), 2010 2nd International Conference*, 2010.

[3] M. Khan, "Context-aware access control for clinical information systems," in *Innovations in Information Technology (IIT), 2012 International Conference*, Tokyo, 2012.

[4] R.-C. Chen, "Development of anti-diabetic drugs ontology for guideline-based clinical drugs recommend system using OWL and SWRL," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference*, Taiwan, 2010.

[5] E. Union, "Collaborative Cardiovascular Research in The European Union's Seventh Framework Programme Activities Supported In The Period 2007-2009," 2009.

[6] G. Zhang, S. Jia and Q. Wang., "Construct ontology-based enterprise information metadata framework," 2010.

[7] R. Sarno, Y. Anistyasari and R. Fitri, Semantic Search, ANDI, 2012.

[8] R. Sarno, L. Yang and C. Bhavsar, "The Agent Matcher architecture applied to power grid transactions," Halifax, 2003.

[9] W. v. d. Aalst, "Process-Aware Information Systems: Lessons to be Learned from Process Mining," *IEEE*, 2009.

[10] W. M. v. d. Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer, 2011.

[11] W. M. v. d. Aalst, K. M. v. Hee, J. M. v. d. Werf and M. Verdonk, "Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor," *IEEE Information Systems*, 2010.

[12] W. v. d. Aalst, H. Reijersa, A. Weijtersa, B. v. Dongena, A. A. d. Medeirosa, M. Songa and H. Verbeeka, "Business process mining: An industrial application," *Elsevier*, 2007.

[13] W. v. d. Aalst, A. Weijters and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, p. 1128–1142, 2004.

[14] W. W. L. M. van der Aalst, "Discovering process models from event logs," *IEEE Transaction on Knowledge and Data Engineering*, 2004.

[15] R. Sarno, Strategi Sukses Bisnis dengan Teknologi Informasi, ITS Press, 2009.

[16] R. Sarno and Heryanti, "Developing Information Technology Policies for Enterprise Resource Planning to Improve Customer Orientation and Service," *International Journal of Computer Science and Network Security*, 2010.

[This page is intentionally left blank]