Knowledge Integration for Diabetes Drugs Ontology

Rung-Ching Chen, Yu-Wen Lo, Bo-Ying Liao, and Cho-Tscan Bau

Department of Information Management, Chaoyang University of Technology,
168, Jifeng E. Rd., Wufeng District, Taichung City 41349, Taiwan ROC
crching@cyut.edu.tw

Abstract. The rising and developing of information technologies has made information overflow. Even in the same topics, a lot of different aspects of knowledge were setup. So, the knowledge integration is one of the important research topics. In this paper, the diabetes drug prescriptions are used as examples to do the knowledge integration which according to American Diabetes Association (ADA), American Journal of Clinical Endocrinology Society (AACE), the Republic of China Diabetes Association and the British National Health Service Bureau (NHS). The system will integrate the four medication diabetes associations to establish knowledge ontologies. The system includes three parts. First, the ontologies pre-processing will calculate the similarity between the ontologies and then find out the correlation between ontologies, Next, the system transfers the ontologies format into Joseki, and finally the user through a graphical user interface to obtain information.

Keywords: Ontology, Ontology Integration, Diabetes Drugs, OWL, SPARQL.

1 Introduction

The knowledge may come from different sources in a particular area such as medical, business and so on. Therefore, the knowledge presentation and structure will be different from various people. So it is important to solve communication barriers from knowledge integration. Then it can achieve to share the knowledge. Through information technology, knowledge can be presented by ontology. At present there are many specific areas of the ontology of knowledge are presented. The main benefits are to find the relationships of knowledge.

The term of ontology is from the philosophy field which refers to "a systematic approach to explain the existence of things in the world" [1]. Ontology can also be used in different fields such as artificial intelligence, semantic web, systems engineering, software engineering and so on. In this paper, we use ontology as a knowledge expression tool. The advantage of ontologies, compared to other data structuring technologies, is the ability to reason upon data. By adding rules and logic it is possible to merge data from different sources, extract and combine data in new ways to produce new information. In general, the elements of ontology include class, attribute and instance [2,3]. Class defines the general things, such as diabetes, diabetes drugs, drugs testing and so on. Attribute: It describes the property of class or the relationship between the concepts. In addition, the relationship between the superclass and the

J.-S. Pan et al. (eds.), *Intelligent Data Analysis and Its Applications, Volume 2*,
Advances in Intelligent Systems and Computing 298,
DOI: 10.1007/978-3-319-07773-4_9,

subclass is also the property. Instance is the entity of concept or class in the ontology. It would inherit all the properties or relation of their class. Nicola [4] proposed the definition of ontology can be based on different conditions to give the different structure and application.

The diabetes drugs are divided into six categories: (1) biguanides, (2) sulfonylurea, (3) DDP4 (DPP4 inhibitor), (4) thiazolidnedione, (5) alpha-glucosidase, (6) meglitinide in Taiwan. Therefore, the six categories of drugs are only considered in this paper[5,6]. In addition, the diabetes drugs are divided into two types: oral hypoglycemic agent (OHA) and insulin. In this paper, only oral hypoglycemic agent to be considered because OHA is convenient to use. Although insulin can effectively control blood glucose, it must be long-term injection. In addition, patients are fear.

Ontology integrations are divided into three kinds: (1) ontology mapping, (2) ontology merging, and (3) ontology alignment. Ontology Mapping is also divided into two parts: One-way mapping and Two-way mapping. One-way mapping refers to ontology A replace ontology B; Two-way mapping refers to ontology A and ontology B can interact with each other. OWL is proposed by W3C and it combines the Darpa Agent Markup Language (DAML) and OIL. The main spirit of OWL is describing the concept of logic[7]. OWL can be divided into three sub-languages: OWL Lite, OWL DL and OWL Full. In this paper, we will use OWL DL to establish diabetes associations and protégé to setup the ontologies of domain knowledge.

In this paper, we will use the domain knowledge of diabetes as an example to do knowledge integration. The main reason of selecting the diabetes domain knowledge is that this disease is one of leading cause of death in the world which is ranked at the top five every year in Taiwan. In addition, diabetes medication knowledge is extensive and it hard to effectively interact and share. This paper will base on four associations included ADA, AACE, Diabetes Association of the Republic of China and NHS to do knowledge integration. We integrated the four medication diabetes associations to establish knowledge ontologies to recommend drugs to the doctors for better treatment for patients.

2 Methodology

Figure 1 shows the flow chart of the drugs recommendation system which divided into five steps:

(1) collect data, (2) create four diabetes associations ontologies, (3) calculate the similarity between the diabetes associations of ontologies, (4) evaluate the relationship between the diabetes associations.

Protégé is used to construct ontologies database in this paper. Protégé is an ontology editor developed by Standford University and the knowledge acquisition software. Protégé use JAVA programming language to develop ontology editor. It provides an interface to users for more flexible applications and access their database and supports RDF (Resource Description Framework). RDF is uesd to describe the relations between web pages and other resources. Protégé mainly has the advantages: easy to learn, free software, supporting a variety of ontologies, supporting multiple store formats, providing complete interface, and providing plug-in program 1. Protégé

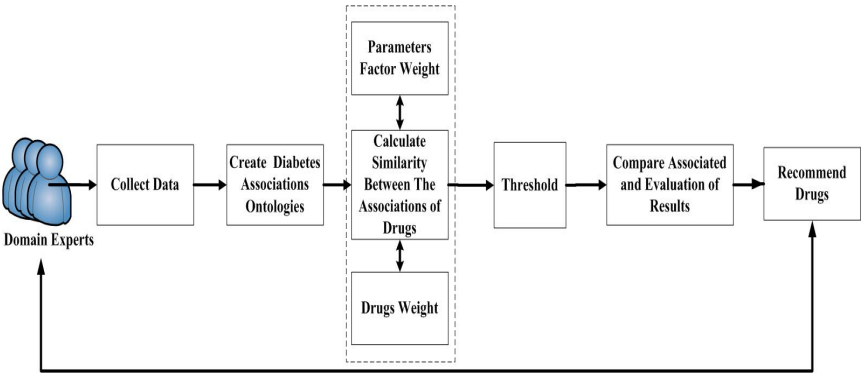

Fig. 1. The flow chart of the drugs recommend system

is an excellent design which allowed many plug-ins. Hence, it has become one of the most widely used ontology editors. In addition, the others more popular ontology editors are such as OntoEdit, OILED, WebODE and so on. In this paper, ontology composition is divided into four parts: (1) class, (2) instance, (3) property, (4) relationship.

2.1 Diabetes Drugs Knowledge Integration

Protégé is used to construct patient ontology and diabetes association ontology. And then we will use Jambalaya to show the various class relationships graph. This paper will use four diabetes association included ADA, AACE, Diabetes Association of the Republic of China and NHS to do knowledge integration. The system can help doctors to diagnose patients in the same situation can give the treatment of diabetes-related knowledge from different countries. It can give more accurate drugs or provide doctors as assistant in the diagnostic process. In this study, four diabetes associations are the main difference between them that they identified the range for glycated hemoglobin (or HbA1c) are different. In this paper, drugs administration divided into two parts: single drugs and combo drugs. Single drug have MET, DPP4, SU, Glinide, TZD and AGI; combo drugs have MET+DPP4, MET+TZD, MET+SU, MET+Glinide, MET+AGI, SU+DPP4, SU+TZD and TZD+DPP4.

The AACE, ADA, Diabetes Association of the Republic of China and NHS ontologies are sequential represented to O_1, O_2, O_3, and O_4.

In addition, the six categories of drugs for diabetes MET, DPP4, SU, Glinide, TZD and AGI are sequential defined as X_{m1}, X_{m2}, X_{m3}, X_{m4}, X_{m5}, and X_{m6}. And then we will define the weight of drugs as W_{m1}, W_{m2}, W_{m3}, W_{m4}, W_{m5}, and W_{m6}. Table 1 shows the symbol of drugs and weights. Each diabetes association of the drugs weight are sum up to one. For example: the sum weight of drugs in O_1 are $W_{m1}+ W_{m2}+ W_{m3},+W_{m4},+W_{m5}+W_{m6}=1$.

Table 1. The definition of drugs and weights

Drug	MET	DPP4	SU	Glinide	TZD	AGI
Symbol	X_{m1}	X_{m2}	X_{m3}	X_{m4}	X_{m5}	X_{m6}
Weight	W_{m1}	W_{m2}	W_{m3}	W_{m4}	W_{m5}	W_{m6}

First, the formula (1) will be used calculate the diabetes association all parameters between the single drug factor values. It's not pure single drug to give to patients if the patient's condition is more serious, the patients will be given the combo drugs. Formula (2) calculates the diabetes association for all parameters between the combo drugs factor values. In this paper, we are focused on single drug and combo drugs.

The parameters have fourteen: blood glucose, liver fat, BMI, waistline, triglycerides (TG), cholesterol (or HDL), hypoglycemia, gastrointestinal dysfunction, renal function indices, liver cirrhosis, liver function, heart failure, fracture risk, and glycated hemoglobin (or HbA1c). The parameter factors are defined as Y_1, Y_2, Y_3, …, Y_i.

$$sum(X_{mi}) = w_{mi} \sum_{j=1}^{n} X_{mi} w_j Y_j \tag{1}$$

$$w_{m1}+w_{m2}+w_{m3}+\cdots+w_{mi} = 1$$

W_{mi} is drug weight, W_j is parameter weight, X_{mi} is drug value, and Y_j is parameter value.

$$sum(X_{mi}) = w_{mi1} \sum_{j=1}^{n} X_{mi_1} w_j Y_j + w_{mi2} \sum_{j=1}^{n} X_{mi_2} w_j Y_j \tag{2}$$

X_{mi1} and X_{mi2} are two of the six categories of drugs, W_{mi1} and W_{mi2} are the weights for the two types of drugs.

After individual diabetes association was evaluated, the system will calculate similarity measurement of four diabetes associations of ontology with the conditions of combination 11 kinds: $sim(O_1,O_2)$, $sim(O_1,O_3)$, $sim(O_1,O_4)$, $sim(O_2,O_3)$, $sim(O_2,O_4)$, $sim(O_3,O_4)$, $sim(O_1,O_2,O_3)$, $sim(O_1,O_2,O_4)$, $sim(O_1,O_3,O_4)$, $sim(O_2,O_3,O_4)$ and $sim(O_1,O_2,O_3,O_4)$. Then, the system calculates the single drug similarity measurement of the two ontologies by the formula (3), and the formula (4) will be used to measure the similarity of combo drugs from two ontologies. In addition, the formula (5) will be used to calculate the single drug and combo drugs similarity measurement form three ontologies, and the formula (6) will be used to calculate the single drug and combo drugs similarity measurement from four ontologies.

$$sim\ (O_{as}, O_{bs}) = \frac{O_{as}(w_{mi} \sum_{j=1}^{n} X_{mi} w_j Y_j)}{O_{bs}(w_{mi} \sum_{j=1}^{n} X_{mi} w_j Y_j)} \tag{3}$$

$sim(O_{as},O_{bs})$ is selected the similarity of single drug from two ontologies.

$$sim\ (O_{ac}, O_{bc}) = \frac{O_{ac}(w_{mi1}\sum_{j=1}^{n} X_{mi_1} w_j Y_j + w_{mi_2}\sum_{j=1}^{n} X_{mi_2} w_j Y_j)}{O_{bc}(w_{mi1}\sum_{j=1}^{n} X_{mi_1} w_j Y_j + w_{mi_2}\sum_{j=1}^{n} X_{mi_2} w_j Y_j)} \quad (4)$$

$sim(O_{ac},O_{bc})$ is selected the similarity of combo drugs from two ontologies .

$$sim(O_a, O_b, O_c) = \frac{sim(O_a, O_b) + sim(O_a, O_c) + sim(O_b, O_c)}{3} \quad (5)$$

$sim(O_a, O_b, O_c)$ is selected the similarity of single drug and combo drugs from three ontologies.

$$sim(O_a, O_b, O_c, O_d) = \frac{sim(O_a, O_b, O_c) + sim(O_a, O_b, O_d) + sim(O_a, O_c, O_d) + sim(O_b, O_c, O_d)}{4} \quad (6)$$

$sim(O_a, O_b, O_c, O_d)$ is selected the similarity of single drug and combo drugs form four ontologies.

2.2 Ontology Mapping Algorithm

In this paper, we will use two algorithms to calculate the similarity of single drug and the combo drugs respectively. They show the results of the ontology mapping. And then they will recommend the availability of drugs. The ontology mapping between the single drug algorithm will input O_1,O_2,O_3, and O_4 parameter factors of drugs; output the single drug

Input: O_1,O_2,O_3, and O_4 parameter factors of drugs
Output: The single drug of ontology mapping results and recommendation drug **Step 1:** extract the ontologies;
Step 2: calculate individual the single drug values for all parameter factors by the formula (1) and $w_{m1}+w_{m2}+w_{m3}+\ldots+w_{mi} = 1$;
Step 3: repeat Step 1 and Step 2, and calculate all the possibilities by the formula $C_2^4 + C_3^4 + C_4^4$, it can find eleven similar combinations;
Step 4: use formula (3) to compare the two of ontologies, formula (5) to compare the three of ontologies and formula (6) to compare the four of ontologies, then $O_{as}(w_{mi}\sum_{j=1}^{n} X_{mi} w_j Y_j)$ and $O_{bs}(w_{mi}\sum_{j=1}^{n} X_{mi} w_j Y_j)$ are random selected ontologies;
Step 5: repeat Step 4, until the completion of the eleven combinations of the calculated;
Step 6: substitute the single drug threshold (T);
Step 7: overtake the threshold value representation the correlation between each other and it's recommended the single drug as the first choice, then it is marked as (**), otherwise, lower than threshold value is recommended as the second choice and it is marked as (*). If the drugs between the ontologies do not be used, this drug is marked as (-);
Step 8: return the single drug of ontology mapping result and recommendation drugs.

Fig. 2. The ontology mapping between the single drug algorithm

of ontology mapping results and recommendation drug. Figure 2 shows the ontology mapping between the single drug algorithms. The ontology mapping between the combo drugs algorithm will input O_1,O_2,O_3, and O_4 parameter factors of drugs; output the combo drugs of ontology mapping results and recommendation drug.

It includes eight steps: (1) extract the ontologies, (2) calculate individual the single drug values for all parameter factors, (3) calculate all the possibilities, (4) use formula (3),(5),(6) to compare all ontologies combination, (5) repeat Step 4, until the completion of the eleven combinations of the calculated, (6) substitute the single drug threshold (T), (7) overtake the threshold value representation the correlation between each other and it's recommended the single drug as the first choice, then it is marked as (**), otherwise, lower than threshold value is recommended as the second choice and it is marked as (*). If the drugs between the ontologies do not be used, this drug is marked as (-), and (8) return the single drug of ontology mapping result and recommendation drugs.

3 The Experiments

The experiments uses Intel(R) Core(TM) Duo CPU P7450 2.13GHz and 4.00 GB RAM. The operation system is Windows 7. Protégé was the platform for constructing ontology. Joseki was used to build the query service. Microsoft Visual Studio 2008 was utilized to develop the user interface system.

This paper will provide recommendation drugs from four diabetes association for doctors. They were based on fourteen different parameter factors to give the drugs in different circumstances. The user interface included three query parts: category query, drugs query and the drugs recommendation. Category query is the overall introduction of the system query and the users can understand the system operation; drugs query is the single drug and combo drugs recommendation results in any parameter factors; the drugs recommendation is shown 11 kinds comparison results.

For drugs query, there are forty-three options. They are included BMI-N (Single Drug), BMI-N (Combo Drugs), Cirrhosis (Single Drug), Cirrhosis (Combo Drugs) and so on. Figure 3 shows the drugs query options. In this paper, three examples are used to explanation for the drugs query. First, it selected the "BMI-N (Single Drug)" to execute. The "BMI-N (Single Drug)" SPARQL query syntax content included three parts: "prefix default: <http://www.owl-ontologies.com/Ontology1305864796.owl#>" is defined to use the prefix, "select ?Parameters ?Drugs" representation wants to query the information, and "where{ {?Parameters a default:BMI} {?Parameters default:single_drug ?Drugs } FILTER regex(str(?Parameters),"BMI-N")}" represents to find BMI parameter, and then find the "default:single_drug" associated with the information between parameters and drugs. In addition, it filters BMI to find subclass BMI-N. Figure 4 shows the "BMI-N (Single Drug)" query result. It displays recommendation drugs of the four diabetes association such as the single drug of BMI-N can use AGI, DPP4 and MET in AACE.

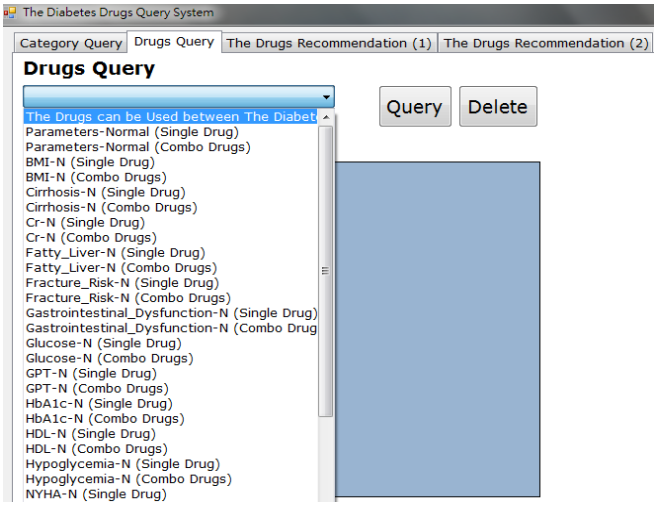

Fig. 3. The drugs query options

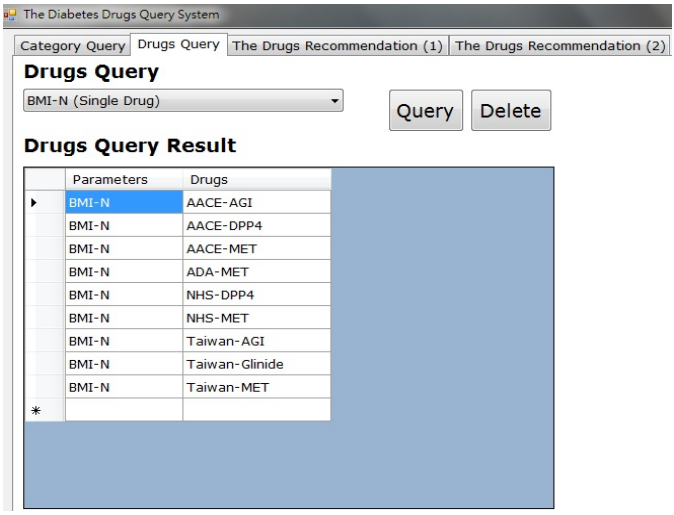

Fig. 4. The "BMI-N (Single Drug)" query result

4 Conclusions and Future Works

Due to the current in the same topics, there are many relevant domains knowledge which come from different countries, domain experts, systems and so on. Therefore, the various related domains knowledge integration is an important work. Ontology can

explain more clearly between the concepts in the particular domain. Hence, we use Protégé to construct domain ontology. In this thesis, the main purpose is to complete the four diabetes drugs knowledge integration and compare similarity between them. In addition, we implement an user interface to help doctors more easier to query diabetes medication knowledge. In this way, our method can help physicians to get more medication knowledge and improve the diagnosis occasion would be more accurate in the drugs administration. In the future works, we will research the triple drugs recommendation and integration. And then we will try to add more medication knowledge among diabetes associations.

Acknowledgement. This work is partially supported by National Science Council, Taiwan, with number: NCS 102-2221-E-324-039.

References

1. Wu, Y.J., Lang, J.S., Shang, F.H.: A Similarity-Based Approach for Ontology Mapping. In: International Conference on Computer Science & Education, pp. 165–169 (2009)
2. Thorsen, K.A.H., Eftestøl, T., Tøssebro, E., Rong, C., Steen, P.A.: Using Ontologies to Integrate and Share Resuscitation Data From Diverse Medical Devices. Resuscitation 80(5), 511–516 (2009)
3. Nicola, G.: Formal Ontology and Information Systems. In: Proceedings of FOIS 1998, Trento, Italy, pp. 3–15. IOS Press (1998)
4. Mario, C., Pietro, H.G., Pierangelo, V.: Using Ontologies for Querying and Analysing Protein-Protein Interaction Data. Procedia Computer Science 1(1), 991–998 (2010)
5. http://www.wretch.cc/blog/f92420f2/10044728
6. http://www.kmu.edu.tw/~kmcj/data/9306/16.htm
7. Corsar, D., Sleeman, D.: Reusing JessTab Rules in Protégé. Knowledge-Based Systems 19(5), 291–297 (2006)