Ontology-based Knowledge Representation for Supporting Medical Decisions

José A. Reyes-Ortiz, Ana L. Jiménez, Jaime Cater, Cesar A. Meléndez, Patricia B. Márquez, Marlon García

Health Digital Systems,
Distrito Federal, Mexico

{alejandro.reyes, ajimenez, jaimecater, cmelendez, pmarquez, marlon.garcia}@saludhds.com.mx

Abstract. The realization of medical diagnoses is the result of decisions taken by doctors, which consists in identifying the signs, symptoms, risk factors and medical background to determine the clinical diagnosis for patients. This process can be performed improperly due to factors such as inaccurate interpretation of information. This paper presents a computational model of representation of medical knowledge to support decision-making task during a medical consultation in order to reduce the chance of misdiagnosis in general medicine. This representation is based on ontologies that provide a mechanism for structuring knowledge to become computer-understandable information, shared by information systems, formalized and using a common vocabulary. Our ontological model is able to infer a list of clinical diagnoses from the data of signs, symptoms, risk factors and medical background.

Keywords: Knowledge representation, clinical diagnosis, ontologies, medical decisions.

1 Introduction

Medical diagnosis is a procedure, which identifies a disease, syndrome or condition of a patient by using several variables among which are relevant the symptoms, risk factors and clinical studies.

The automation of this process is a challenge for computational systems. In this area doctors generate the final diagnoses manually in systems based in the Electronic Health Record (EHR), which selects the final diagnosis from the international catalog of diseases called ICD10 [1]. This process is tedious, very time consuming and can be generated incorrectly due to misuse of the doctor-patient relationship, interrogation and misdirected clinical misinterpretation of medical semiotics or due to doctor's physical wear. It has been estimated that the percentage of medical errors during hospital care is among 3.5% to 16.6% [2]. These problems reflect the importance of

generating information systems that enable medical support during the decision making process to provide a final diagnosis. A computational model that represents medical knowledge should support such systems.

This paper puts forward a computational model based on ontology representation, which is used to support decision-making in clinical diagnostic generation from input data as semiotics, medical history and risk factors. This model is able to infer a list of likely clinical diagnoses relying on a mechanism of inference (rules) on the stored information. The model is based on ontologies due to its ability to represent and share knowledge in an automated manner, the use of a shared vocabulary in the field of medical diagnostics and the ability to formalize knowledge in ontological language accepted as standard. In addition, ontologies are able to infer new knowledge by relying on inference machines, so they are a novel technology to knowledge representation in order to improve diagnosis process from certain medical conditions, such as semiotics, medical history and risk factors.

The rest of this paper is organized as follows. Section 2 presents the elements involved in the process of obtaining a medical diagnosis and the theoretical foundations of knowledge representation based on ontologies. Section 3 presents the work related to the task of obtaining automatic or semi-automatic clinical diagnoses, as well as state of the art work of carrying out knowledge representation through ontologies. The knowledge representation model based on ontologies proposed in this paper is presented in section 4. Section 5 presents the implementation of the proposed ontology, which sets in all classes, relationships and instances of our ontology. Finally, Section 6 presents the conclusions and Section 7 provides the corresponding acknowledgments.

2 Theoretical Framework

The work presented in this article is based on a knowledge representation model to support the work of medical decision-making in clinical diagnostics. Therefore, in this section we present the elements needed to make a medical diagnosis and the theoretical foundations of knowledge representation based on ontologies.

2.1 Medical Diagnosis

Medical diagnosis is a procedure, which identifies a disease, syndrome, or any health condition. This process is carried out using various criteria, such as: symptoms, risk factors, sex and age of patients, laboratory, desk studies and vital signs, such as temperature, blood pleasure, pulse rate and respiration rate.

The most relevant criteria for diagnosis are symptoms with anatomy and evolution, risk factors such as addictions, family-inherited background including the relative (family member) involved. Symptoms are identified by diagnostic printing process, which is clinical reasoning in order to explain the condition. To consolidate a diagnosis it is necessary to identify internal and external risk factors. Once the final

diagnosis is made, the doctor proceeds with the treatment, control and monitoring of the disease.

In the area of clinical care, a doctor assigned to the consultation performs the process described above. This process can be done incorrectly, resulting from various factors, the most common documented mistakes are: improper use of the doctor-patient relationship, misdirected interrogation and clinical misinterpretation of semiotics [3].

Computer systems can support this process. However, knowledge representation models are necessary to organize the information needed to make the diagnosis in order to make this information manageable by a machine. This model can be understood by a machine to perform inference on the information and get the diagnoses that meet the symptoms and risk factors.

Our work proposes a model of representation of the ontology-based information to make it manageable for an expert system that supports decision-making in medical diagnosis.

2.2 Knowledge Representation

Knowledge representation is an Artificial Intelligence area research which aims to facilitate the inference of new knowledge from the elements represented symbolically. The knowledge representation also involves an analysis of the proper and efficient cognitive reasoning to perform a certain task [4].

The proper use of given information leads to a set of facts or assumptions in a knowledge domain. This information makes logical sense that helps to formalize the semantics of how reasoning should work.

Knowledge representation can be carried out with the support of the ontologies, which function as the means to accomplish this task. Ontologies become the most viable means to accomplish this representation because of its benefits and features [5] and [6]:

1. They offer a specification using a conceptual model.
2. The model is processed by a computer.
3. Allows sharing knowledge using a common vocabulary.
4. The vocabulary used is accepted by a group of domain experts

In our approach, ontologies are proposed to represent medical information necessary to apply inference rules and get a set of likely clinical diagnoses. These ontologies are our model of knowledge representation.

3 Related Work

In the reviewed literature, there are studies that have supported the process of carrying out a medical diagnosis; in this respect computational approaches have been proposed for the generation of probable diagnoses, which have been addressed semi-

automatically or automatically. In this regard there are two types of approaches: a) the learning-based statistical approaches such as [7] presented an update of the algorithm C4.5 to "learn" from examples and determine the diagnosis, a statistical approach using Support Vector Machines (SVM) algorithm is discussed in [8], b) finally, approaches based on rules generated manually by experts [9] and [10]. In this respect, there are no studies that support ontology to represent medical knowledge and perform clinical diagnostic process.

The knowledge representation based on ontology is a technique that has been used in domains such as bioinformatics and molecular biology in order to capture existing processes in these domains [11], a protein ontology is proposed in [12] to support computer systems in biology in general providing a proper interpretation of the data, and to model the relationships in proteins, [13] proposes a model of an ontology-based annotation for disease phenotypes, which facilitates the discovery of new phenotype-genotype relationships among species.

The indications of generating computer systems that support medical diagnoses are based on statistical approaches, leaving aside the semantic interpretation of the input data. Meanwhile, the representation of knowledge is an area that has been applied extensively in medicine, specifically in bioinformatics. However, research on ontology-supported medical diagnosis is a challenging task that has no investigation involved.

Knowledge representation of diagnosis has been addressed from other perspectives as probabilistic and logic based approaches. Hence, [14] proposes an approach that uses Bayesian networks in order to argue and deal with the uncertainty problem of fault diagnosis well, the Bayesian network structure is established according to the cause and effect sequence of faults and symptoms, which are the only medical conditions considered in the model of authors; [15] applies the concept of a fuzzy set as knowledge representation to improve the decision making process.

Therefore an attempt is made by authors to design and develop a diagnosis system. The system developed is evaluated using a simple set of symptoms that is added to clinical data in determining diabetes and its severity. [16] offers an approach to create a new medical knowledge representation model, bases on the use of probabilistic theory. Their work start from the description, realized by an expert of the medical knowledge describing the relation between symptoms and diagnoses the proposed approach consists on building a probabilistic model including the Medical Knowledge Base. Moreover, the proposed approach integrates several probabilistic reasoning mechanisms based on the considered knowledge.

Most of the diagnosis systems are designed to possess the clinical data and symptoms associated with a specific disease as knowledge base or they use few or simple medial conditions. Our work presents a novel approach for knowledge representation based on ontologies that unifies the terminology used in diagnosis, facilitates diagnostic prediction and understanding of diagnoses. We consider several medical conditions that are used to determine a clinical diagnosis.

4 Proposed Knowledge Representation Model

This paper puts forward an ontological model to represent medical information, which facilitates obtaining a diagnosis. The information is modeled based on the concepts (disease, symptoms, anatomy, risk factors and background) and the relationships between them (symptom-anatomy, disease-symptom, etc.).

The proposed ontological model is presented in a modular fashion, which has been divided into the following sections: diseases (age group and gender); symptoms with anatomy, intensity and evolution, and risk factors.

A diagnosis must be an instance of the disease concept and it is governed by the International Catalogue of Diseases ICD-10 [1]. Therefore, three sub-ontologies are created that shape the complete model: *Diseases, Symptoms and RiskFactors*. Figure 1 shows the complete model of the Diagnosis Ontology and their relationships between sub-ontologies.

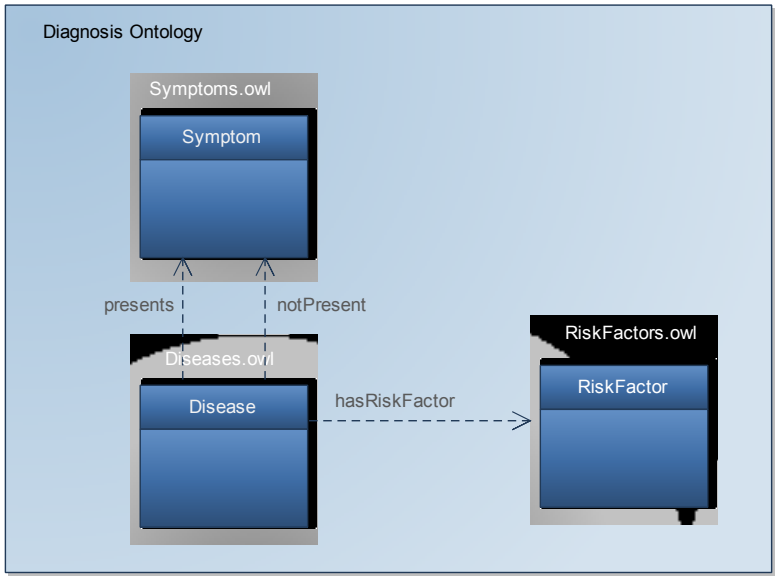

Fig. 1. Diagnosis ontology model

4.1 Diseases

Diseases Ontology has the *Disease* class and has 21 subclasses: *Infectious, Neoplasia, BloodDisease, Endocrine, MentalDisorder, NervousSystemDisease, EyeDisease, EarDisease, CirculatorySystemDisease, RespiratorySystemDisease, DigestiveTractDisease, DermatologicDisease, OsteomusclarSystemDisease, GenitourinarySystemDisease, Pregnancy-Birth, PerinatalInfection,*

CongenitalMalformation, AbnormalClinicalFinding, Trauma-Poisoning and MorbidExtremeCauses.

Disease class is related to class *Gender, AgeGroup, Symptom* and *RiskFactor*. Figure 2 shows *Diseases* Ontology with some examples of subclasses and their corresponding relationships.

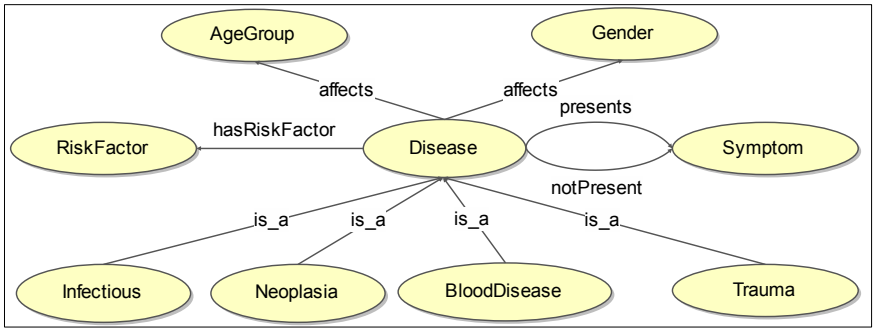

Fig. 2. Diseases ontology

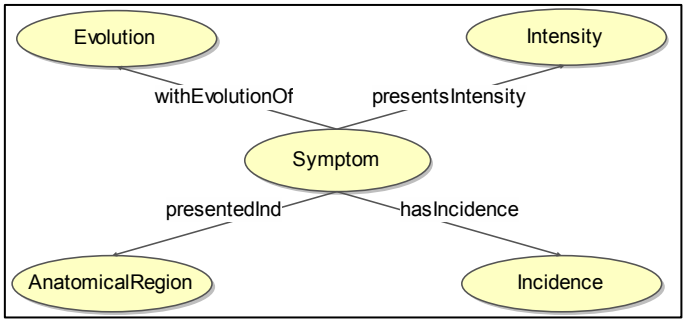

Fig. 3. Symptoms ontology

The *presents* and *notPresent* relations express the absence or presence of certain disease symptoms. On the other hand, the *affects* relationship represents both gender and age group likely to be diagnosed with a certain disease. Finally, the *hasRiskFactor* relationship is used to represent circumstances that allow the existence of a diagnosis.

The *Disease* class instances are the list of possible diagnoses, they belong to the ICD-10 and some examples of *Infectious* subclass instances are paratyphoid *fever*, *pulmonary tuberculosis* and *neonatal tetanus*.

4.2 Symptoms

Symptoms ontology shapes knowledge presented in the disease symptoms. The main class of this ontology is *Symptom* which has four relationships *presentedIn*,

withEvolutionOf, *presentsIntensity* and *hasIncidence* towards *AnatomicalRegion, Evolution*, *Intensity*, and *Incidence* classes respectively. Figure 3 shows the ontology created to represent the relationship between the symptoms and the diseases.

Examples of Symptom class instances are: *headache*, *pruritus*, and *asthenia*. Meanwhile anatomical regions express human body parts where certain symptom happens, e.g. *stomach*, *legs*, *temporal region of the head*. The intensity, evolution and incidence are relations of the symptoms that are characteristic for certain diseases, which help to differentiate and make a wise decision.

4.3 Risk Factors

A risk factor is a condition or component that can cause a disease to be diagnosed. The main class of this ontology is *RiskFactor*, which has three subclasses *PersonalBackground*, *Addiction* and *FamilyInheritedBackground*. Figure 4 shows the ontology to represent the risk factors that indicate the presence of a certain disease diagnosed.

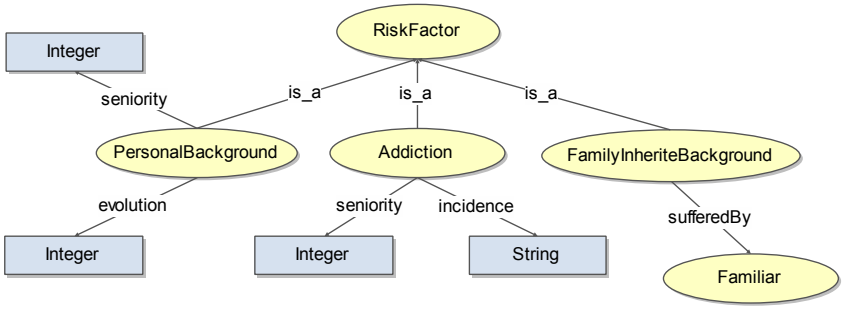

Fig. 4. Risk factor onrology

The *PersonalBackground* class has two properties of object relationships seniority and evolution which are Integer type. Meanwhile, the Addiction class has two properties of object relationships: seniority, which is Integer type; and incidence, which is String type. Finally, the *FamilyInheritedBackground* class has a relationship to Family class called *sufferedBy*, which is used to express the relative (*father*, *mother*, *uncle*, *grandfather/grandmother*, *brother*/*sister*) who suffer/s a disease.

5 Implementation

The implementation of ontology was made relying on the framework based on knowledge and ontology editor called Protégé [17]. The ontology was encoded using the Web Ontology Language in its version 2.0, which provides classes, properties, individuals and data type values [18]. This language is stored as documents under the Semantic Web standards.

The global ontology called Diagnostic Ontology contains 78 classes or concepts, 23 relations of data type, 29 object type relations. Obtaining a diagnosis is based on the mapping of the input data (patient data, history, symptoms, etc.) with the information stored in the ontology for diseases. We are currently working on this phase.

The information stored in the ontology by classes, relations, individuals and types of data is validated and accepted by a group of medical experts that ensure a common vocabulary in the area of diagnostics.

The fact of using an ontological model to represent information enables the exchange of knowledge, the use of a common vocabulary in applications that are supported by this ontology and reuse of knowledge represented.

6 Conclusions

In this paper we have presented a model that integrates ontological concepts (classes), real entities (entities or individuals) and interactions between classes (relations). These elements add the meaning and the information necessary to perform a diagnosis. This is supporting the process of decision-making by the doctor in order to obtain a final diagnosis.

The proposed model turns medical knowledge in a semantic structure used by a machine for various tasks, such as automatic retrieval of information and inferences. The ontological model presented focuses on three sub-ontologies *Disease*, *Symptoms* and *RiskFactors*. In the *Diseases* Ontology, semantic structured information of diagnoses is represented and is the core of the global model. *Symptoms* ontology is used to represent the information that leads to disease. *RiskFactors* ontology is a model that expresses the necessary conditions for a disease to be diagnosed.

The ontology created is used to support the task of obtaining a list of probable diagnoses. However, in order to obtain this list is not enough with the ontological model, it is essential a set of inference rules to extract new knowledge from existing data. This task leads a research and further work.

We are currently working on these rules to achieve the automatic generation of a list of likely diagnoses from input data. This generation will support medical decision-making, specifically in the generation of the final diagnosis by a doctor. Our approach does not propose the full replacement of a physician in the clinical diagnosis process for patients. However, it proposes a support for physician in decision-making in order to improve the clinical diagnosis process.

Our ontological model can be extended to consider new medical conditions such as family history, gynecological history, laboratory studies, vital signs (temperature, blood pressure and pulse rate), among others. It will be closer to all medical considerations in making a diagnosis manual process.

Our proposed ontology cannot only be used to support decision-making in diagnoses; it can also be used for the task of extracting information on diseases. In addition, it can be used to support a question answer system about the symptoms of diseases, such as: *What are the symptoms of X?* or *What risk factors are conducive to*

X?. This task also involves a great challenge and Natural Language Processing techniques are needed in order to understand the question.

Acknowledgements. This project was funded through the National Council of Science and Technology of Mexico (CONACYT) under project No. 164 487.

References

1. Organización Mundial de la Salud: Clasificación Internacional de Enfermedades, décima versión CIE-10. Informe de la Conferencia Internacional para la Décima Revisión de la Clasificación Internacional de Enfermedades (1998)
2. Martínez, C. M.: Errores médicos en la práctica clínica, del paradigma biologicista al paradigma médico social. Revista Cubana Salud Pública 32(1). Ciudad de La Habana (2006)
3. Díaz, J., Gallego, B., León, A.: El diagnóstico médico: bases y procedimientos. Revista Cubana de Medicina General Integral. Revista Cubana Medicina General Integral 22(1), Ciudad de La Habana (2006)
4. Davis, R., Shrobe, H., Szolovits, P.: What Is a Knowledge Representation? Association for the Advancement of Artificial Intelligence 14(1), pp. 17−33 (1993)
5. Gruber T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology 43(5−6), pp. 907−928. USA (1993)
6. Borst, W.: Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Universiteit Twente, PhD thesis, Enschede (1997)
7. Alexopoulos, E., Dounias, G. D., Vemmos, K.: Medical Diagnosis of Stroke using Inductive Machine Learning. In: Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course on Artificial Intelligence ACAI99, pp. 91−101, Chania, Greece (1999)
8. Shashikant, U. G., Ghatol A. A.: Heart Disease Diagnosis Using Machine Learning Algorithm. In: Proceedings of the International Conference on Information Systems Design and Intelligent Applications. Advances in Intelligent and Soft Computing 132, pp. 217−225, India (2012)
9. Drastal, G. A., Kulikowski, C. A.: Knowledge-Based Acquisition of Rules for Medical Diagnosis. Journal of Medical Systems 6(5), pp. 433−445, Springer, Netherlands (1982)
10. Jianwei, X., Ke, X.: A novel extracting medical diagnosis rules based on rough sets. In: The 2009 IEEE International Conference on Granular Computing, pp. 608−611, IEEE Press, Nanchang, China (2009)
11. Steven, R., Goble, C.A., Bechhofer, S.: Ontology-based knowledge representation for bioinformatics. Brief Bioinform 1(4), pp. 398−414, United Kingdom (2000)
12. Chang, E., Sidhu, A., Dillon, T., Sidhu, B.: Ontology-based Knowledge Representation for Protein Data. In: Dillon, T. and Yu, X and Chang, Elizabeth (ed.), INDIN 2005: 3rd International Conference on Industrial Informatics, Frontier Technologies for the Future of Industry and Business. IEEE Press, pp. 535−539. Australia (2005)
13. Washington, N. L., Haendel, M. A., Mungall, C. J., Ashburner, M., Westerfield, M., Lewis, S. E.: Linking Human Diseases to Animal Models Using Ontology-Based Phenotype Annotation. PLoS Biol. 2009 Nov; 7(11) USA (2009)

14. Baoping, C., Yonghong, L., Qian, F., Yunwei, Z., Zengkai, L., Shilin, Y., Renjie, J.: Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network. Applied Energy, vol. 114, pp. 1−9. (2014)
15. Anouncia, M. S., Maddona, C. L., Jeevitha, P., Nandhini, R. T.: Design of a Diabetic Diagnosis System Using Rough Sets. Cybernetics and Information Technologies, vol. 13 (3), pp. 124−169. (2013)
16. Alsun, M., Lecornu, L., Solaiman, B.: Possibilistic Medical Knowledge Representation Model. J. Adv. Inf. Fusion, vol. 7(2), pp. 101−113. (2012)
17. Tudorache, T., Noy, N., Tu, S., Musen, M.: Supporting Collaborative Ontology Development in Protégé. The Semantic Web – International Semantic Web Conference, vol. 5318, pp. 17−32. (2010).

Smith, M., Horrocks, I., Krötzsch, B. G.: OWL 2 Web Ontology Language: Conformance (Second Edition). Eds. W3C Recommendation. Latest version available at http://www.w3.org/TR/owl2-conformance/. (2012)