Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-108

Ontology Driven Decision Support Systems for Medical Diagnosis

An interactive form for consultation in patients with plasma cell disease

Valéry DONFACK GUEFACK[a,1], Valérie BERTAUD GOUNOT[a], Régis DUVAUFERRIER[a], Annabel BOURDE[a] John MORELLI[b], Jérémy LASBLEIZ[a]
[a] *Unité Inserm U936, Faculté de Médecine, Rennes 1 University, France*
[b] *Department of Radiology, Scott and White Memorial Hospital and Clinic, Texas A&M Health Sciences Center*

Abstract. Multiple myeloma (MM) is a malignant disorder characterized by the monoclonal proliferation of B cell derived plasma cells in the bone marrow. The diagnosis depends on the identification of abnormal monoclonal marrow plasma cells, monoclonal protein in the serum or urine, evidence of end-organ damage, and a clinical picture consistent with MM. The distinction between MM stages—monoclonal gammopathy of undetermined significance or indolent myeloma—is critical in guiding therapy. This paper describes how to produce ontology-driven semiological rules base (SRB) and a consultation form to aid in the diagnosis of plasma cells diseases. We have extracted the MM sub-ontology from the NCI Thesaurus. Using Protégé 3.4.2 and owl1, criteria in the literature for the diagnosis and staging of MM have been added to the ontology. All quantitative parameters have been transformed to a qualitative format. A formal description of MM variants and stages has been given. The obtained ontology has been checked by a reasoner and instantiated to obtain a SRB. The form created has been tested and evaluated utilizing 63 clinical medical reports. The likelihood for a disease being the correct diagnosis is determined by computing a ratio. The resulting tool is relevant for MM diagnosis and staging.

Keywords. NCI Thésaurus, OWL, biomedical Ontology, medical semiology, qualitative knowledge, quantitative knowledge

Introduction

Multiple myeloma (MM) is a malignant disorder characterized by the monoclonal proliferation of B cell derived plasma cells in the bone marrow [1]. The distinction between monoclonal gammopathy of undetermined significance (MGUS) and indolent myeloma can be difficult, especially if the disease is clinically occult [2]. The diagnosis of MM depends on the identification of abnormal monoclonal plasma cells in the bone marrow, monoclonal protein in the serum or urine, evidence of end-organ damage, and a clinical picture consistent with MM. MM has been classified by various schema, whereby certain criteria are utilized to determine diagnosis, staging, risk stratification, and appropriate therapy [1]. The NCI Thesaurus (NCIT) is the major ontological

[1] Corresponding Author, e-mail: valerydonfack@hotmail.com

resource in the cancer domain [3]. However, the descriptions of MM, MGUS and other related disorders in the NCIT do not correspond to operational definitions in the literature [1]. We assume that ontology with generic tools (an ontology editor and reasoner) can be helpful for the medical diagnosis of these entities if diagnostic criteria of diseases are well-formalized [4], and this study expands upon the results of a previous work [5]. In that study, we showed how to create an ontology-driven rules base for expert systems. The objectives of this study are to operationalize an ontology for plasma cell diseases using established criteria from the medical literature [1] to construct an interactive form aiding in the diagnosis of plasma cell disorders. This will then be applied to sixty-three clinical medical reports. To that end [5], (1) we will extract a sub ontology of the NCIT (plasma cell disease), (2) add, update, and reorganize concepts to comply with the most recent literature [6], (3) check consistency of this ontology with a Pellet reasoner [5], (4) add necessary rules and instantiate the ontology according to [5], (5) save the content of the ontology in a MYSQL database and (6) attempt to validate the system utilizing sixty-three clinical medical reports.

1. Materials and methods

1.1. Building the ontology

The construction of our ontology begins with the NCIT ontology [3]. **OWL API** [7], **Jess**, and **SWRL** are useful for applications for concept extraction, concept reorganization, etc. We utilized Protege 3.4.2 as the ontology editor, and the Pellet 1.5.2 reasoner for checking the ontology. In addition to the four steps elaborated on in reference [5], we reorganized and updated [8] any vague or quantitative concepts and stored a set of semiological rules in a relational database. The name of these new concepts carried the prefix of «**U936_**».

1.2. Concept reorganization

MGUS is defined as a pre-malignant condition in references [2] with only 1% of cases progressing to MM. However in the NCIT, MGUS is defined as a neoplasm akin to MM. The ontology has been revised to reflect MGUS as a pre-malignant condition.

1.3. Concept addition

The MM staging system utilized in the NCIT is incomplete, consisting of an older version of the Salmon and Durie system [1] rather than the Salmon and Durie PLUS system. Furthermore, separate categories of symptomatic myeloma, Waldenstrom‘s disease, and MGUS [1] are not included. In reference [1] MGUS is defined as having serum monoclonal protein < 30 g/l, clonal bone marrow < 10%, as well as possessing other, less clearly factors such as the absence of hypercalcemia, renal insufficiency, anemia and bone lesions (CRAB) [1]. In the NCIT, the serum monoclonal protein is not quantified and factors relating to the CRAB criteria are not represented.

1.4. Medical reports

To evaluate and validate the system we utilized sixty-three clinical medical reports.

1.5. Storing semiological rules in a relational database, building the form and querying

To facilitate ease of access to the system, the semiological rule base was stored in a relational database (MYSQL) for validation and evaluation.

2. Results

After extraction, we obtained a sub ontology of plasma cell diseases with 277 class, 16 roles and 822 restrictions [5]. The reorganization of «**May_Have**» roles was performed as in reference [5].

2.1. Concept reorganization and addition

In compliance with the operational definitions of (MM) in references [1], MGUS was linked to pre-malignant states. The missing concepts like « symptomatic myeloma », « Waldenstrom's disease », and the Salmon and Durie Plus classification [9] were added. All quantitative criteria (e.g: ***Serum monoclonal protein < 3 g/dl)*** were transformed into qualitative parameters, as only the presence or absence of signs and symptoms was evaluated: *U936_IgA_Serum_Measurement*

U936_IgA_Serum_Measurement_More_Than_3_GPerDl
U936_IgA_Serum_Measurement_Under_3_GPerDl

The qualitative parameters which were less clearly defined (e.g. "Hypercalcemia") were characterized in compliance with last outcomes in MM:

Serum_Calcium_Measurement
U936_Serum_Calcium_Measurement_More_Than_2.6_Or_3_MmolPerL
U936_Serum_Calcium_Measurement_Under_Or_Equal_2.6_Or_3_MmolPerL

2.2. Adding new roles and adding an SWRL rule layer for abductive reasoning and management of an « Excludes » role

Four new roles were added for management of the four SWRL rules as in reference [5].

2.3. Creating the prototypical cases [5]

SWRL reasons based on instances. In this process, the ontology's hierarchy is conserved [10]. According to the subsumption relation [11], all the individuals of the subclasses of a given class are also its individuals.

2.4. Storing of semiological rules in a relational database, creating the form

The ontology content was stored in MYSQL relational database and the form was created. Database queries were performed via this form. The signs are displayed by category on the form [5].

2.5. Validation and evaluation

The form was validated by using the definitions of MM in [1] and was evaluated on sixty three real medical reports. An input form was utilized (http://nautilus.univ-rennes1.fr:8080/AppDiagnosticRatioWithDetail/Auth.do). The semilogical rules in the MYSQL relational database were used to provide an accounting for all possible and excluded diagnoses. All possible diagnoses proposed at physicians are included under the category of possible diagnoses. For each possible diagnosis, the ratio of the number of signs which are present in a given case to the number of total signs evaluated is also displayed. Thus, the set of all potential diagnoses can be ranked by their relevance [5]. In addition, missing or unobserved signs are also displayed. (Table 1)

Table 1 – Example of patient case

Finding		**Diagnostic**	
Present	**Absent**	**Possibilities**	**Excluded diagnoses**
Hemoglobin >10g/dl, IgA<3g/dl, IgG< 5g/dl, Calcium <3 mmol/l, UPEP<4g/24h	Bone pain, X ray findings, Splenomegaly	Indolent myeloma, Stage I MM, Stage IA MM, Stage IB MM	MGUS, IgM MGUS, Waldenstrom, Stage II, Stage II A et B MM, MM, Solitary osseous plasmocytoma, smoldering myeloma, plasma cell leukemia, non-secretory MM, Poems syndrome

3. Discussion

Previous works have described the creation of ontology-driven rule bases for expert systems [5]. Our study demonstrates how to generate a semilogical rule base and a form from an ontology written in OWL for consultation in the diagnosis of MM. In this work, we deal with only signs which are present or absent with all quantitative data being transformed into qualitative parameters. This study of MM shows that the NCIT's definitions of MM is not in compliance with the operational definitions of MM in the literature [1]. In the NCIT definitions, some concepts are not represented, missing, or defined vaguely. In reality, although an ontology enables knowledge representation and sharing, reasoning, and guidance for decision-making, it does not enable medical diagnosis [12]. As per references [1,2], the ontology must be updated or operationalized [8] before medical diagnoses can be rendered. In addition to illustrating previous transformations in the development of this system [5], in the present work we proceeded in (1) reorganizing concepts of MM to be in compliance with operational definitions set forth in the current medical literature [1], (2) representing quantitative and vague concepts as in reference [8], and adding any missing concepts per the definitions set forth in references [1,13]. In comparison to reference [5], the creation of prototypical cases is done while preserving hierarchy [10] according to the subsumption relation [11]: all the individuals belonging to a subclasses of a given class are also individuals of that class. To determine the likelihood of a given diagnosis to be correct, the number of observed signs that suggest the presence of a given diagnosis is divided by the total number of total signs characterizing the disease. Querying the database in conjunction with the results form provides relevant information regarding the knowledge base. The system shows signs which are present, absent, present and imply/exclude a given diagnosis, and signs that are missing thus implying/excluding a

diagnosis. Any signs which are apparently missing or not observed can be explored further by the physician. Validation and evaluation on the 63 actual medical reports demonstrate that 90% of cases to be well-classified and 10% not well-classified by the system. Well-classified case is one of the possibilities, it may be the highest ranking possibility or another in the list. The 10% of misclassified cases are attributable to features not represented in the ontology. The knowledge linked to these cases can be integrated into the knowledge base by a learning process.

In conclusion, the approach outlined herein demonstrates how to operationalize ontology for medical diagnosis. Quantitative and vague diagnostic criteria are dealt with by using qualitative concepts in their description. We think that, if we use quantitative knowledge with threshold values to describe them, as in [8], it will avoid necessitating the use of several parameters to describe several states of the same concept. The classification of diseases by the described method is made by assuming that all signs have similar weight in rendering a diagnosis. This is unlikely to be reflective of reality. We would optimally like to weight the value of the observed/unobserved signs by utilizing quantitative data such as sensitivity and specificity [14]. This knowledge can be theoretically integrated directly in ontology by including a data type property [4].

References

[1] Rajkumar SV, Multiple myeloma: 2011 update on diagnosis, risk-stratification, and management, American Journal of Hematology, vol. 86, no. 1, pp. 57-65, Jan. 2011.
[2] Decaux O et al., Monoclonal gammopathies of undetermined significance and their progression over time. Retrospective study of 190 patients, La presse médicale 2006, 35(7-8):1143-1150
[3] Fragoso G, de Coronado S, Haber M, Hartel F, and Wright L, Overview and utilization of the NCI thesaurus.,Comparative and functional genomics Jan. 2004; 5(8):648-654
[4] Bertaud-Gounot V, Duvauferrier R, and Burgun A, "Ontology and medical diagnosis., Informatics for health & social care Oct. 2011:1-11
[5] Bertaud-Gounot V, Donfack Guefack V, Lasbleiz J, Bourdé A, and Duvauferrier R, Creating an ontology driven rules base for an expert system for medical diagnosis, Studies in Health Technology and Informatics, 169:714-718 2011.
[6] Noy NF, de Coronado S, Solbrig H, Fragoso G, Hartel WF, and Musen MA, Representing the NCI Thesaurus in OWL DL: Modeling tools help modeling languages, Applied ontology, vol. 3, no. 3, pp. 173-190, 2008.
[7] Wikipedia, The OWL API," 2003. [Online]. Available: http://owlapi.sourceforge.net/.
[8] Peelen L, Klein MCA, Schlobach S, Keizer NFD, and Peek N, Analyzing Differences in Operational Disease Definitions Using Ontological Modeling, Proceed AIME, 297-302, 2007.
[9] Kyle RA and Rajkumar S, Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma, Leukemia 2009, 23(1):3-9
[10] Parsia B and Sirin E, Pellet : An OWL DL Reasoner, Artificial Intelligence 2000.
[11] Shiffman RN and Greenes RA, Rule set reduction using augmented decision table and semantic subsumption techniques: application to cholesterol guidelines., Proceedings Symposium on Computer Applications in Medical Care, 339-343, Jan. 1992.
[12] Burgun A, Bodenreider O, and Jacquelinet C, Issues in the Classification of Disease Instances with Ontologies, Studies in Health Technology and Informatics 2005, 116:695-700
[13] Durie BGM, The role of anatomic and functional staging in myeloma: description of Durie/Salmon plus staging system,European journal of cancer Jul. 2006, 42(11): 1539-1543
[14] McGee S, Simplifying likelihood ratios, Journal of general internal medicine, Aug. 2002; 17(8):646-649