Creation Of Definitions for Ontologies: a case study in the Leukemia Domain

Amanda Damasceno de Souza[1] Maurício Barcellos Almeida[2] Joaquim Caetano de Aguirre Neto[3]

[1]School of Information Science, Federal University of Minas Gerais, Belo Horizonte, Brazil, amanda@ufmg.br
[2]Department of Theory and Management of Information, Federal University of Minas Gerais, Belo Horizonte, Brazil, mba@ufmg.br
[3]*Santa Casa de Belo Horizonte* Hospital, Minas Gerais, Belo Horizonte, Brazil, caetanoaguirre@hotmail.com

Abstract. The creation of the definitions it is an important stage of the activity of ontologies construction, insofar as the definitions provides the understanding of the meaning of classes. However, creating definitions is a complex and tiresome task. This study is part of an ongoing research that analyses some fundamental principles with the aim of formulating textual and formal definitions to be used in ontologies. The context of such analysis is a project of knowledge organization within the biomedical domain. The goal is to establish methodological guidelines for formulating the definitions in biomedical ontologies. In general, people building ontologies do not make use of consistent rules for the correct formulation of definitions, which, we believe, make our study a relevant initiative. As partial results, we present a list of topics that corresponds to the aforementioned methodological guidelines.

Keywords: Definitions. Biomedical Ontologies. Leukemia.

1 Introduction

In the context of the development of new information technologies, there are great potential for the use of ontologies for organizing medical information. Ontologies have been largely applied in the biomedical field, which demands semantic tools to better represent the large amount of medical entities and terms [1,2]. Indeed, the use of ontologies is an alternative that has been receiving an increased amount of attention [3]. One step in building ontologies is the formulation of well-formulated definitions. Understanding how to create definitions is very important in order to organize concepts and terms for purposes of information representation and retrieval. This study aims to systematize the process of the creating definitions in the biomedical ontologies. In order to do this, we present a study case in the leukemia domain. Leukemia

has having a strong impact in modern society due to the low rates of patients' survival. In addition, leukemia is a complex disease due of the phenotypic heterogeneity. The class called *Acute Myeloid Leukemia* (AML) corresponds to a set of heterogeneous diseases related to the clonality and chromosomal alterations [4].

Ontologies should provide clear and coherent definitions of the structures that are found in reality [5]. In order to make definitions understandable for computers, one has to create textual definitions and then translate them to some form of logic. An ontological hierarchy depends on the specification of properties that defines the essence of entities. This essence provides the basis on which such entities can be grouped together and distinguished one from another. The main role of definitions in ontologies is to emphasize those properties, as well as satisfying the need of transitive inheritance in hierarchies. The position of a class in a hierarchy can contribute to the understanding of its meaning [5].

In this paper, we discuss some ontological principles in the scope of construction of a large biomedical ontology (*Blood Ontology* – BLO [6]). We seek to formulate definitions for Leukemia within the cancer domain. One might claim that this effort does not present any research contribution or novelty. However, we believe in the relevance of our initiative, insofar as biomedical vocabularies and medical texts in general exhibit several sorts of mistakes in formulating definitions [14].

2 Methods

The terminological sample for our case study was taken from BLO. We aim to define a range of classes bellow AML, which contains 24 subclasses (Fig. 1). We also intend to define other hematological neoplasms, namely: a) Myelodysplastic syndrome (containing 5 subclasses); b) Myeloproliferative neoplasm (containing 11 subclasses).

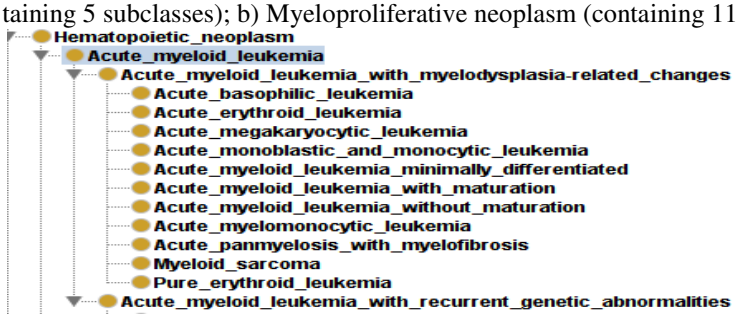

Fig.1.- 24 classes of AML.
Source: BLO in Protegé, Almeida *et al.* [6]

We have systematized criteria for the natural language and formal logic language definitions based on the best practices proposed in the literature [5] [7] [8] [9] [10] [11] [12]. The steps of formulating textual definitions are part of our preliminaries results. In order to reach our preliminary findings we made use of a list of topics (from a to g):

a) to understand the meaning of the term using more than one sources
b) to establish the higher genus in the context of use of the term
c) to establish the essential characteristic of the entity
d) to formulate the definition in the form *S = Def. G which Ds*, where "G" stands for genus (the parent of S); and "S" stands for species
e) to verify whether the definition is a statement of necessary and sufficient conditions
f) to verify whether the definition is non-circular
g) to verify the existence of multiple-inheritance and try to eliminate it

The first class of our hierarchy, as well as its definition, came directly from BLO: "*An hematopoietic neoplasm is a hematologic malignancy which occurs in blood-forming tissues*". The second class was defined as acute myeloid leukemia (AML). Those definitions are the starting point of searching the essential feature of AML and its inheritance. Our next step in the context of the project is to formulate formal definitions using a logical language.

3 Preliminary Results

As we have mentioned before, some features of a class can be obtained by checking its inheritance. So, an AML received characteristics from the correspondent upper class, namely, hematopoietic neoplasm, which has characteristics in common with other classes in the hierarchy of BLO for blood cancers. The distinction between AML and other leukemia types is the myeloid cell lineage. Using the hierarchy of AML in BLO is possible to define the first relation of AML as a subsumption <is_a> relation, which connects a class to another one <class, class>. So, *acute myeloid leukemia is_a hematopoietic neoplasm*. Among other possible relations to define ALM, one can highlight the of the relation of derivation *c <derives_from> c1*, for example: *Acute Myeloid Leukemia derives-from hematopoietic stem cell*. Those relations are based on two material continuants [1], each one distinct of each other. Derivation is a relation between instances, where a simple continuant creates a plurality of other continuants. Some other examples of definitions based on class-class sort of relation are: <has_a> as in: *Acute Myeloid Leukemia has_a Clonal Disorder*; and *Acute Myeloid Leukemia has_a myeloid (monocytic) lineage*. Using the class-class relation <Located_in> relation, one can found: *Acute Myeloid Leukemia Located_in Blood* [13]. We used the definition of AML to illustrate the process of formulating textual definition on leukemia domain: Df = `A leukemia that occurs when a hematopoietic stem cell undergoes malignant transformation into a primitive, differentiated cell with abnormal longevity and with abnormal proliferation of myeloid cells lineage`. The main contribution of our approach is to emphasize the need of adopting some rules for creating definitions in ontologies. In general, people building ontologies don't follow any guidelines to create definitions.

4 Final Remarks

We present part of an ongoing project within Information Science field. We show our preliminary and partial results in defining a range of biomedical terms. This initial stage is presented with the aim of emphasizing the need of some guidelines or even a list of topics to formulate proper definitions. This will helps one, for example, to understand that the nature of things can be different (continuants and occurrrents), as well as other required distinctions, for example, that relations among instances are different of relations among classes. So, we expect that in using our list of topics, one will be able to build better ontologies and provide advances in the development of expert medical systems. In reason of space limitations, we don't present any example here, but we intend to do this in future papers.

References

1. Spear AD. Ontology for the Twenty First Century: An Introduction with Recommendations. *Institute for Formal Ontology and Medical Information Science (IFOMIS)*, Saarbrücken, Germany .University at Buffalo, New York, U.S.A. (2006). http://ifomis.buffalo.edu/bfo/documents/manual.pdf.
2. Washington N, Lewis S. Ontologies: Scientific Data Sharing Made Easy. *Nature Education*.1(3), 5 (2008)
3. Almeida MB, Barbosa RR. Ontologies in knowledge management support - a case study. *J Am Soc. Inf. Sci. Technol.* 60(10), 2032-47 (2009)
4. Wernig G, Gilliland G. Pathobiology of acute myeloid leukemia.In: Hoffman R, *et al.* (ed.). *Hematology: basic principle and practice*. 5th.ed. Churchill Livingstone: Elsevier, pp.921-932 (2009)
5. Michael J, Mejino JL Jr, Rosse C. The role of definitions in biomedical concept representation.*Proc. AMIA Symp*. 463-7(2001)
6. 6.Almeida MB, *et al*. The Blood Ontology: an ontology in the domain of hematology. *Proceedings of the International Conference of Biomedical Ontologies*, (2011). http://ceur-ws.org/Vol-833/.
7. Köhler J, *et al*. Quality control for terms and definitions in ontologies and taxonomies.*BMC Bioinformatics*. 7,212 (2006)
8. Smith B, *et al*. Relations in biomedical ontologies.*Genome Biol*. 6(5),R46 (2005)
9. Seppälä S, Ruttenberg R. Survey on Defining Practices in Ontologies: Report in the International Workshop on Definitions in Ontologies .In: *ICBO 2013*, Montreal, Quebec, Canada,(2013). http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013_surveyreport_seppalaruttenberg.pdf.
10. Petrova A, *et al*. Formalizing biomedical concepts from textual definitions.*J Biomed Semantics*. 6,22-(2015)
11. Seppälä S, Schreiber Y, Ruttenberg, A. Textual and logical definitions in ontologies. In: *Proceedings of DIKR 2014, IWOOD 2014, CEUR Workshop Proceedings*, Houston, TX, USA, (1309)35-41 (2014).https://seljaseppala.files.wordpress.com/2014/11/seppc3a4lc3a42014a.pdf.

12. Tsatsaronis G, *et al*. Learning Formal Definitions for Biomedical Concepts.In: Srinivas, K,Jupp S. (eds) *Proceedings of the 10th OWL*: Experiences and Directions Workshop (OWLED), (2013). https://ddll.inf.tu-dresden.de/web/LATPub509/en.
13. Schulz S, Kumar A, Bittner T. Biomedical ontologies: what part-of is and isn't.J. *Biomed. Inform.*39(3),350-61 (2006)
14. Liss PE, Aspevall O, Karlsson D, Forsum U.Terms used to describe urinary tract infections--the importance of conceptual clarification. *APMIS*. Volume 111, n. 2, 291–299, (2003).