Textual Definitions in the Leukemia Domain:

Methodological Guidelines for Biomedical Ontologies

Amanda Damasceno de Souza[a] Maurício Barcellos Almeida[b]

[a]*School of Information Science, Federal University of Minas Gerais, Belo Horizonte, Brazil*

[b]*Advisor ˉ Department of Theory and Management of Information - Federal University of Minas Gerais - Belo Horizonte, Brazil*

Abstract

The present study, part of an ongoing research, analyses some fundamental principles that aims to formulate textual and formal definitions to be used in ontologies. The goal of our study is to establish methodological guidelines for the formulation of definitions in biomedical ontologies, which are not so common in the field of Library & Information Science.

Keywords:

Definitions; Biomedical Ontologies.

Introduction

In this study, we aim to systematize the process of creating definitions. The step we have taken can be considered general steps for creating definition, but we apply them in an ontology in the blood cancer domain, particularly in the leukemia domain. We discuss applied ontology principles in the construction of a fragment of an biomedical ontology, seeking to formulate textual and formal definitions for that field of cancer. This study is part of construction of the *Blood Ontology* (BLO) [1].

Ontologies should provide clear, coherent definitions of structures found in reality written in a form that is computer understandable. Consistency in definitions requires a unifying viewpoint. In order to make these definitions understandable for computers, one has to create textual definitions and then translate them to some formal logic. A good definition is a fundamental aspect for the understanding of concepts and for the integration languages and makes them compatible [2].

Methods

The terminological sample was taken from BLO, an ontology about blood. We start defining kind of AML, which has 24 classes as shown in Figure 1. We intend also to define another hematological neoplasms, namely, both Myelodysplastic syndrome with 5 classes and Myeloproliferative neoplasm with 11 classes.
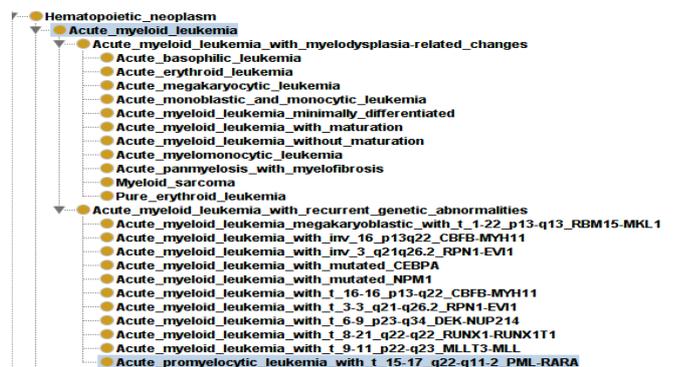

Figure 1 - 24 classes of AML.
Source: BLO in Protegé, Almeida *et al.* [9]

We have systematized criteria for natural language and formal logic language definitions based on the best practices proposed by Köhler *et al.* [3], Smith *et al.* [4],Seppäla and Ruttenberg [5], to mention but a few. At this moment, our research don´t bring a complete complete literarure review, and some references was adopted for testing the creation of a set of guidelines. The steps in formulating textual definitions are part of our preliminaries results, organized in list as follows: a) separate the term; b) get a preliminary definition of the meaning of the term in any source; c) establish the higher genus in the context of use of the term; d) establish the

essential characteristic distinguishing the genus from species; e) formulate the first version of the definition in the form S = Def. G which Ds, where "G" stands for genus, which is the parent of S; and "S" stands for species; f) check if the definition is a statement of necessary and sufficient conditions; g) check principle of non- circularity; h) check the principle of multiple-inheritance.

Preliminary Results

AML received the characteristics from the correspondent upper class, namely, *hematopoietic neoplasm*, which has characteristics in common with other classes in the hierarchy of BLO blood cancers. The difference between AML and other leukemia types is the myeloid cell lineage. Through the hierarchy of AML in BLO it is possible to define the first formal relation of AML as an *is_a* relation *<class, class>*, where *acute myeloid leukemia is_a hematopoietic neoplasm*. We used the definition of AML to illustrate the process of formulating 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.* We don´t believe we are providing a contribution to the state of at of ontological engineering only trying some cases in our leukemia ontology. The main contribution of our approach is the it is educational, insofar as people building ontologies in the field of Library & Information Science, in general, don't have guidelines to build good definitions.

Among the possible relations to define formally the ALM, one can highlighted the relation *c derives_from c1*, for example: *Acute Myeloid Leukemia derives-from hematopoietic stem cell.* These relations are based on two material continuants, each one distinct of each other. Derivation is a relation between instances, in which a simple continuant creates a plurality of another continuants. An example in biomedical domain where one can found cell derivation is when *promyelocytic* cells derives from *myeloblastic* cells [4].

Conlcusion

This is an ongoing research project within Information Science field, in which we present preliminary and partial results. We presented the initial stage our research, in the hope that soon we are able to define other types of leukemia.

References

 [1] Almeida MB, Proietti ABFC, Ai J, Smith B. The Blood Ontology: an ontology in the domain of hematology. *Proceedings of the International Conference of Biomedical Ontologies*, 2011, US. Available at: http://ceur-ws.org/Vol-833/. Access April 2014.

[2] Campos MLA. The role of the definitions in ontology research. *Perspect Ciênc Inf* 2010: 15(1):220-38. [Portuguese].

[3] Köhler J, Munn K, Rüegg A, Skusa A, Smith B. Quality control for terms and definitions in ontologies and taxonomies.*BMC Bioinformatics*. 2006: 7:212.

[4] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies.*Genome Biol* 2005: 6(5):R46.

 [5] Seppälä S, Ruttenberg R. Survey on Defining Practices in Ontologies : Report. in preparation of the International Workshop on Definitions in Ontologies (DO 2013). IN: *International Conference On Biomedical Ontology* (ICBO 2013), 4, July 2013, Montreal, Quebec, Canada, 2013. Available at: http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013_surveyreport_seppalaruttenberg.pdf. Access Nov 2014.

Address for correspondence

Amanda Damasceno de Souza, Escola de Ciência da Informação, Universidade Federal de Minas Gerais.

Rua José Batista filho, 300, Centro – Sarzedo – MG – Brazil, 32450-000, amanda@ufmg.br.