



XV Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB 2014

GT 11 – Informação em saúde

PRINCÍPIOS METODOLÓGICOS PARA CRIAÇÃO DE DEFINIÇÕES DE TERMOS ESPECIALIZADOS EM ONTOLOGIAS BIOMÉDICAS: UM ESTUDO SOBRE LEUCEMIAS NO DOMÍNIO DO CÂNCER

METHODOLOGICAL PRINCIPLES FOR THE CREATION OF DEFINITIONS OF SPECIALIZED TERMS IN BIOMEDICAL ONTOLOGIES: A STUDY ABOUT LEUKEMIA IN THE CANCER DOMAIN

Amanda Damasceno de Souza; Maurício Barcellos Almeida

Modalidade da apresentação: Comunicação oral

Resumo: A criação de definições é uma importante etapa na construção de ontologias, uma vez que elas propiciam entendimento semântico sobre um termo. Isso é essencial para localizar um termo na hierarquia e para estabelecer relações com outros termos. Criar definições formais em ontologias é uma tarefa árdua, complexa e que consome tempo, de forma que qualquer apoio sistemático pode ser de grande valia para o cientista da informação. O presente estudo, parte de pesquisa em andamento, analisa diversas formas de definir e seus princípios fundamentais, com vistas à formulação de definições textuais e formais em ontologias. O contexto para tal análise consiste de um projeto de pesquisa que contempla a construção de um *sistema de organização do conhecimento* no domínio do câncer de sangue, especificamente sobre as Leucemias Mieloides Agudas. Buscam-se aqui determinar princípios metodológicos para a formulação de definições em ontologias biomédicas, bem como a validação dessas definições com especialistas. Como resultado parcial, apresentam-se passos correspondentes a tais princípios metodológicos. Conclui-se que metodologias propostas para construção de ontologias ainda não contemplam diretrizes sólidas para a formulação de definições, o que torna este estudo relevante.

Palavras-chave: Definições. Ontologias Biomédicas. Leucemia.

Abstract: The creation of definitions is an important phase of the activity of ontologies construction, insofar as the definitions provide semantic understanding about terms. This is essential to properly locate the term in the hierarchy and to establish relations with other terms. Creating formal definitions in ontologies is a hard, complex and tiresome task. Thus, any systematic support can be of great value to the information scientist. The present study, part of an ongoing research, analyses several ways of defining terms, as well as the fundamental principles of defining, with the aim of formulating textual and formal definitions for ontologies. The context of such analysis consists of a research project that includes the construction of a knowledge organization system in the domain of blood cancer, particularly

about acute myeloid leukemia. We aim to provide methodological principles for the formulation of definitions in biomedical ontologies, as well as the validation of those definitions with experts. As partial results, we present a list of topics that corresponds those methodological principles. We conclude that methodologies for ontology construction do not include consistent guidelines for the correct formulation of definitions, which make this study a relevant initiative.

Keywords: Definitions. Biomedical Ontologies. Leukemia.

1 INTRODUÇÃO

A busca pela definição para um termo está relacionada ao processo de aprendizagem, a compreensão do termo, seus significados e usos. A definição é assim uma fonte de aprendizado. Entretanto, utilizar ou entender uma linguagem não significa ser capaz de formular a definição de um termo (SWARTZ, 2010). Isso ocorre porque o processo de criação de definições envolve decisões que devem ser baseadas em critérios pré-estabelecidos e bem documentados (USCHOLD, 1996).

Ontologias biomédicas são importantes recursos uma vez que registram e organizam o conhecimento sobre biomedicina, integram informações provenientes de diferentes sistemas e servem como base para o desenvolvimento de sistemas especializados. Entre as ontologias biomédicas pode-se citar, por exemplo, a *Blood Ontology* (BLO)¹, a *Gene Ontology* (GO)², a *Foundational Model of Anatomy Ontology* (FMA)³, dentre outras. As metodologias para construção de ontologias têm como importante etapa a criação de definições para os termos. Definições de termos médicos podem ser encontradas em dicionários médicos ou em vocabulários controlados como o MeSH (*Medical Subject Headings*)⁴ e o *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED-CT)⁵, dentre outros. Entretanto, as definições encontradas nestes tipos de recursos nem sempre não são adequadas a ontologias biomédicas. Isso resulta na necessidade de criar definições segundo princípios ontológicos.

Definições apresentam um papel importante na construção de ontologias consistentes, as quais devem funcionar como um sistema integrado e propiciar entendimento semântico (CAMPOS, 2010). A presença de definições bem fundamentadas é um dos critérios para se avaliar a qualidade das ontologias. A OBO *Foundry Library*, que segundo Smith *et al.* (2007) é um repositório para desenvolvimento e registro de ontologias sobre ciências da vida, exige que uma ontologia cumpra alguns requisitos para ser incluída no repositório. O primeiro requisito diz respeito às definições textuais: como devem ser formuladas de forma a assegurar a precisão da ontologia ao mesmo tempo em que permitam compreensão humana. Cabe citar que existem iniciativas para aperfeiçoar definições textuais e formais em grandes ontologias

¹ <http://mba.eci.ufmg.br/BLO/>

² <http://www.geneontology.org/>

³ <http://sig.biostr.washington.edu/projects/fm/index.html>

⁴ <http://www.nlm.nih.gov/mesh/>

⁵ http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

já existentes, no sentido de corrigir erros de circularidade e intangibilidade (SMITH, *et al.*; 2005; KÖHLER *et al.*, 2006).

O presente trabalho é parte de pesquisa em andamento cujo objetivo é propor princípios metodológicos para a criação de definições em ontologias biomédicas. A pesquisa vem sendo desenvolvida no contexto de um projeto cujo objetivo é construir um *Knowledge Organization System* (KOS) no domínio do sangue (ALMEIDA *et al.* 2011). Apresenta-se aqui a parte do KOS relativa ao câncer de sangue, especificamente, às leucemias mieloides agudas.

No âmbito da Ciência da Informação (CI), o presente estudo se justifica pela importância de estudos sobre como criar definições e pelas similaridades verificadas entre ontologias e outros KOS amplamente consagrados na CI. Estudos sobre definições já vêm sendo feitos na CI, como por exemplo, os estudos definitórios citados por Campos (2010) na elaboração de ontologias consistentes e, mesmo os clássicos estudos de Dalhberg (1978a,b), que através da Teoria do Conceito possibilitou a determinação e o entendimento do conceito, para fins de representação e recuperação (CAMPOS, 2005; 2010). Na verdade, a CI já faz uso, diretamente ou indiretamente, de muitos princípios ontológicos que podem ser úteis para representar e para recuperar informação (ALMEIDA, 2013). Espera-se contribuir com princípios sistemáticos para a criação de definições em ontologias, bem como fornecer subsídios para profissionais atuando com ontologias no âmbito da CI.

No âmbito da biomedicina, a pesquisa se justifica pela relevância do tema para a sociedade. Apenas no ano de 2012 estima-se 350 mil novos casos e 265 mil óbitos causados por essa doença em todo mundo. Para o Brasil, o Instituto Nacional de Câncer (INCA) estimou para 2014 cerca de 9370 novos casos de leucemias. Para o Estado de Minas Gerais são esperados 910 casos novos de leucemias e 140 casos em Belo Horizonte (INCA, 2014). Dessa forma, qualquer apoio automatizado para organização e recuperação da informação como o proporcionado pelas ontologias, torna-se relevante para lidar com a grande massa de dados produzida em biomedicina.

2 BACKGROUND: LEUCEMIAS MIELOIDES AGUDAS (LMA)

A leucemia tem um papel complexo na sociedade moderna devido aos altos índices de incidência combinados a uma baixa sobrevida dos pacientes, além de representar um dos cânceres que mais acomete crianças (KAMPEN, 2012). De acordo com o *National Cancer Institute* (NCI, 2013) as leucemias são agrupadas em quatro tipos mais frequentes de acordo

com sua evolução, em crônica (mais insidiosa) ou aguda (de apresentação súbita e com progressão rápida), e pelo tipo de glóbulo branco que é afetado:

1. Leucemia mieloide aguda (LMA): mais frequente em adultos, ocorre quando os blastos leucêmicos acumulam-se na medula óssea e sangue.
2. Leucemia linfóide aguda (LLA): tipo mais comum em crianças.
3. Leucemia mieloide crônica (LMC): acometendo principalmente adultos, ocorre quando o sangue apresenta um aumento do número de células brancas.
4. Leucemia linfóide crônica (LLC): em geral, acomete pessoas com mais de 55 anos.

Leucemia mieloide aguda (LMA) refere-se a um grupo de doenças heterogêneas, com respeito à clonalidade, alterações cromossômicas e resposta ao tratamento. A avaliação clínica e o prognóstico de pacientes com a LMA vêm mudando drasticamente ao longo da última década. Estudos citogenéticos, moleculares e imunológicos têm contribuído para o entendimento da patogênese e prognóstico da LMA (NAJFELD, 2009; WERNIG; GILLILAND, 2009).

A pesquisa sobre a criação de definições em ontologias biomédicas é conduzido no domínio das neoplasias hematológicas, e se restringe especificamente ao domínio das leucemias mieloides agudas (LMA).

2 DEFINIÇÕES

No âmbito da Ciência da Informação (CI), Campos (2005) explica que ainda nos anos 60, Dahlberg desenvolveu a Teoria do Conceito destinada a elaboração de tesouros, com a possibilidade de utilizar princípios de elaboração de terminologias para o domínio das linguagens documentárias de abordagem alfabética. A Teoria do Conceito possibilitou uma compreensão mais sólida do conceito direcionada à representação e à recuperação da informação, ao desenvolver princípios para estabelecer as relações entre conceitos, fornecendo elementos para a criação de definições consistentes. Em Dahlberg (1978a,b), o conceito é formado por três elementos: o referente, as características e a forma verbal. Hjørland (2009) analisa que conceitos são onipresentes e penetrante no campo da Biblioteconomia e Ciência da Informação (BCI), apresentando um papel importante na recuperação da informação. O autor afirma que teorias como a do conceito, estão relacionadas a organização do conhecimento, para o desenvolvimento de sistemas de classificação, taxonomias, tesouros e ontologias. As diferentes formas que se apresentam as teorias sobre

conceitos têm implicações na forma como CI investiga seus temas centrais. Neste contexto, Soergel *et al.* (2004) afirmam que “O maior desafio da recuperação de informação é a identificação do conceito em um domínio específico de interesse!”⁶

Entre as varias formas de definir termos, Robinson⁷ (1950 citado por SWARTZ, 2010) explica definição como todo recurso que pode ser usado para ensinar outra pessoa a como utilizar um termo linguístico. A norma ISO 704 – *Terminology work – Principles and methods* (2000, p.vii), conceitua “definição” como algo que descreve um conceito. Mesmo que existam diferentes acepções para o termo “definição”, em geral as definições apresentam varias características e funções, como por exemplo:

- a) Aumentar e enriquecer vocabulário
- b) Introduzir o significado e uso de novas palavras para as pessoas
- c) Eliminar alguns tipos de ambiguidade
- d) Reduzir a imprecisão
- e) Solucionar problemas epistemológicos

No discurso comum há vários objetos possíveis de definição e, conseqüentemente, diversas formas de se definir uma coisa. Sobre os tipos de definições, Gupta (2008) e Swartz (2010) citam os seguintes:

- a) Definições reais e nominais: definições reais se referem às características que constituem o objeto, as qualidades, as propriedades das quais este objeto dependem; as definições nominais se referem a ideias abstratas do objeto e explicam o significado de um termo.
- b) Definições lexicográficas ou lexicais: dicionários e glossários, que explicam o significado de um termo restrito a certo sentido específico.
- c) Definições ostensivas: referem-se ao significado do termo através de exemplo e contexto onde este é utilizado.
- d) Definições persuasivas: forma de definição que se propõe a descrever o verdadeiro ou significado socialmente aceito de um termo; relacionam-se a termos que visam alterar direitos ou deveres e envolvem noções imprecisas, como liberdade ou democracia.

⁶ The biggest challenge in information retrieval is concept identification in a specific domain of interest!

⁷ ROBINSON, R. **Definition**. Oxford: Clarendon Press, 1950.

- e) Definições estipulativas: especificam e restringem como um termo que apresenta vários significados deve ser usado no contexto específico em questão.
- f) Definições descritivas: referem-se ao significado do uso já existente para o termo.

Na teoria de definições o símbolo a ser definido é chamado de *definiendum* e o conjunto de símbolos usados para explicar o significado do *definiendum* é chamado de *definiens*. Onde (... X ...) é chamado de *definiendum* da definição, e a expressão do lado direito (=Df -----) é o seu *definiens* (a definição em si). O símbolo padrão utilizado para representar a definição é representado por “=Df”. Assim a fórmula do termo definido (como x se relaciona com =df) pode ser representada da seguinte forma:

$$\dots X \dots = \text{Df} \text{-----}.$$

Na lógica de definições é necessário seguir alguns critérios como, por exemplo, o critério do conservadorismo, o qual postula que uma definição não deve permitir estabelecer novas reivindicações, novas afirmações. A NORMA ISO 704 (2000) aborda que a definição deve definir o termo em uma intensão e extensão única:

- a) Intensão/conotação: indica o conceito superordenado, imediatamente superior seguido pela característica que distingue um termo de outros.
- b) Extensão/denotação/referente: trata-se de uma lista de termos completa em que os termos subordinados podem ser esclarecidos por definições intensionais.

2.1 Definições em ontologias: princípios aristotélicos

Campos (2010, p.222) argumenta que “no caso das ontologias, as definições propiciam a possibilidade de compatibilização semântica, pois descrevem o conteúdo semântico de um termo”. Segundo a autora as metodologias de construção de ontologias não apresentam diretrizes satisfatórias na identificação de conceitos, dos seus tipos de relacionamento e de como criar definições para estes conceitos. Prover definições de qualidade se configura como uma tarefa desafiadora e que consome tempo.

Ao formular definições para ontologia é necessário seguir alguns princípios como a *herança única*, o qual postula que as definições em ontologias devem estar dispostas em forma hierárquica de características semelhantes, de forma que os termos inferiores recebam por herança os conceitos dos termos superiores. O princípio de herança nas ontologias se vale

das características comuns entre os termos, quando a definição do termo anterior enriquece a definição do termo posterior. A definição de um termo na ontologia será incompleta sem as heranças de seus pais (MICHAEL; MEJINO JUNIOR; ROSSE, 2001, p.463).

Outro princípio importante é o da *não circularidade*, que postula que, um termo não deve ser usado para definir ele mesmo (KÖHLER *et al.*, 2006). A importância de definir termos na ontologia seguindo o princípio de não circularidade diz respeito a prover definições que ofereçam informações além das inerentes ao próprio termo definido (ver exemplo de circularidade no QUADRO 1). Köhler *et al.* (2006) afirmam que definições devem ser escritas de forma clara para que se compreenda o real significado do termo, devem ser concisas em frases curtas e completas.

QUADRO 1 – Exemplo de Circularidade da Gene Ontology (GO)

id:	GO:0042270
term:	<i>Protection from natural killer cell mediated cytotoxicity</i>
definition:	<i>The process of protecting a cell from cytotoxicity by natural killer cells</i>

Fonte: Köhler *et al.* (2006).

O princípio da *intangibilidade* sugere que, ao formular definições em ontologias, evite-se a linguagem figurativa ou obscura (KÖHLER *et al.* 2006). Esta regra postula que ao formular a definição deve-se adotar termos que sejam mais inteligíveis do que o que está sendo definido (vide exemplo no QUADRO 2). De acordo com o princípio de intangibilidade, cada termo definido deve ter atender aos padrões básicos de compreensão. Para que a definição seja inteligível, ela precisa ser compreensível sem que sejam necessárias leituras prévias ou consultas a fontes de informação especializada. Além disso, também é preciso evitar o uso de terminologia técnica.

QUADRO 2 - Exemplo de intangibilidade na GO

id:	GO:0050566
term:	<i>asparaginyl-tRNA synthase (glutamine-hydrolyzing) activity</i>
definition:	<i>Catalysis Cyc:6.3.5.6-RXN</i>

Fonte: Köhler *et al.* (2006, p.4).

Outra questão relevante no contexto das definições criadas de acordo com princípios aristotélicos dizer respeito às condições necessárias e suficientes. Na verdade, uma definição é uma declaração de condições necessárias e suficientes (SMITH, 2013). Por exemplo: sendo A é uma condição necessária para ser um B, todo B é um A; sendo A é uma condição suficiente para ser um B, todo A é um B; e assim pode-se definir um A como uma coisa que satisfaz B. Isto ocorre porque A é um termo mais difícil de ser compreendido do que B. As condições necessárias e suficientes de uma definição estão relacionadas ao conceito de intensão. Este princípio é explicado através da lógica, conforme citado por Swartz (2010):

QUADRO 3 – Condições suficientes e necessárias

“x é uma condição suficiente para y” =_{df} “a presença (/existência / verdade) de x garante a presença (/existência / verdade) de y”

“x é uma condição necessárias para y” =_{df} “a ausência (/não-existência /falsidade) de x garante a ausência (/não-existência /falsidade) de u=y”

*ser um quadrado é condição suficiente para ser retangular
 ser retangular é condição necessárias par ser um quadrado
 ser uma mãe é uma condição suficiente para ser uma mulher
 ser uma mulher é condição necessária para ser uma mãe
 ter quatro lados é uma condição suficiente para ter um numero par de lados
 ter um número par de lados é uma condição necessária para ter quatro lados*

Fonte: Swartz (2010).

Swartz (2010) explica, com os exemplos citados, que as relações *é uma condição suficiente* e *é uma condição necessária* são implicações inversas, ou seja, se x é uma condição suficiente para y, então y é uma condição necessária para x.

Há um grande número de tipos de condições suficientes, e para cada tipo há um correspondente de condição necessária. Em certos casos, algumas condições são logicamente suficientes para outras; em certos casos, algumas condições serão logicamente necessárias, mas podem não ser suficientes. Assim se x é uma condição lógica suficiente para y, então y é uma condição lógica necessária para x. Entretanto algumas condições são necessárias para alguns casos, sem ser logicamente suficientes, mas sim ser uma condição física suficiente para outro caso (SWARTZ, 2010).

Köhler *et al.* (2006, p.2) recomendam a utilização das seguintes regras básicas para formulações de definições textuais, baseadas nos princípios aristotélicos:

- a) Concentrar-se em características essenciais,
- b) Evitar circularidade,
- c) Capturar a extensão correta,
- d) Evitar linguagem figurativa ou obscura,
- e) Evitar uso de negativos.

2.2 Definições em ontologias biomédicas sobre leucemias mieloides agudas

A presente seção descreve a pesquisa em andamento, no contexto de desenvolvimento de uma ontologia sobre sangue (BLO), onde se busca formular princípios metodológicos para a criação de definições para termos em ontologias.

Ao realizar definições em linguagem textual (ou natural) sobre as leucemias, o primeiro termo a ser definido foi *Hematopoietic neoplasm*, o qual definido pela BLO como “*an hematopoietic neoplasm is a hematologic malignancy which occurs in blood-forming tissues*”⁸. O segundo termo a ser definido foi leucemia mieloide aguda (LMA). Essas definições são o ponto de partida para a definição da característica essencial da LMA e suas heranças.

Como exemplo de definições em linguagem textual, obtém-se os termos *Leucemia* (*Leukemia*) e leucemia aguda (*Acute Leukemia*) no Ontobee (2014a):

Leukemia (Term IRI: http://purl.obolibrary.org/obo/DOID_1240):

Definition: *A cancer that affects the blood or bone marrow characterized by an abnormal proliferation of blood cells. [database_cross_reference:*

url: <http://en.wikipedia.org/wiki/Leukemia>]

(ONTOBEE, 2014a).

Acute leukemia (Term IRI: http://purl.obolibrary.org/obo/DOID_12603):

Definition: *A leukemia that occurs when a hematopoietic stem cell undergoes malignant transformation into a primitive, undifferentiated cell with abnormal longevity. These lymphocytes (acute lymphocytic leukemia [ALL]) or myeloid cells (acute myelocytic leukemia [AML]) proliferate abnormally, replacing normal marrow tissue and hematopoietic cells and inducing anemia, thrombocytopenia, and granulocytopenia. Because they are bloodborne, they can infiltrate various organs and sites, including the liver, spleen, lymph nodes, CNS, kidneys, and gonads. [database_cross_reference:*

url: <http://www.merck.com/mmpe/sec11/ch142/ch142b.html>].

(ONTOBEE, 2014a).

A Ontobee é uma fonte de informação projetada para ontologias que visa facilitar o compartilhamento, a consulta, a integração e a análise de ontologias. Trata-se de um repositório de dados para a maioria das ontologias registradas no OBO *Foundry Library* (ONTOBEE, 2014b; XIANG *et al.*, 2011).

⁸ O termo e a definição foram mantidos em inglês de acordo com a concepção original da BLO.

Ao buscar a definição de LMA em outras fontes especializadas, foram encontradas as seguintes opções:

- a. WINTROBE: *Acute myeloid leukemia*. =Df. *refer to a group of marrow-based neoplasms that have clinical similarities and distinct morphologic, immunophenotypic, cytogenetic, and molecular features* (ARBER; COUSA, 2013).
- b. NATIONAL CANCER INSTITUTE (NCI): *Acute myeloid leukemia*. =Df. *is a type of cancer in which the bone marrow makes abnormal myeloblasts (a type of white blood cell), red blood cells, or platelets* (NATIONAL CANCER INSTITUTE, 2014).
- c. MESH: *Acute myeloid leukemia* =Df. *Form of leukemia characterized by an uncontrolled proliferation of the myeloid lineage and their precursors (MYELOID PROGENITOR CELLS) in the bone marrow and other sites* (MEDICAL SUBJECT HEADINGS, 2004).
- d. MAYO CLINIC: *Acute myeloid leukemia* =Df. *is a cancer of the blood and bone marrow — the spongy tissue inside bones where blood cells are made* (MAYO CLINIC, 2014).

Nos exemplos acima, ao se definir LMA foi evitada a circularidade. Entretanto, não é clara qual é a característica principal da LMA, assim não é claro se os princípios ontológicos foram seguidos. Nem todas as definições apresentaram a característica essencial da LMA, ou seja, linhagem mieloide. A definição **a** é uma definição geral laboratorial. A definição **b** cita uma característica essencial “*makes abnormal myeloblasts*”, mas cita também características não essenciais como “*red blood cells, or platelets*”, que são gerais para outros tipos de leucemias. A definição **c** foi à melhor e citou a característica essencial: *myeloid lineage*, já a definição **d** é muito geral, voltada ao paciente por apresentar uma conotação não científica.

2.3 Definições formais da leucemia mieloide aguda

Nas ontologias biomédicas, a formulação de definições formais (ou lógicas) ocorre por meio de relações, uma vez que as relações desempenham um papel importante em conectar as classes (SMITH *et al.*; 2005). Ao formular definições, o princípio de herança é fundamental, sendo este o primeiro a ser aplicado. Neste contexto a relação *is-a (é-um)* fornece a semântica formal para que as características de heranças sejam repassadas aos termos na hierarquia. Assim, relações são formas de proporcionar o suporte às classes na hierarquia. Definir uma relação *is-a* entre dois termos, significa que o primeiro é um subtipo do segundo, por exemplo: *Acute myeloid leukemia is-a hematopoietic neoplasm*, significa que *Acute myeloid leukemia* é um subtipo de *hematopoietic neoplasm*.

Essas relações são criadas através da definição de classes e instâncias, na qual temos três tipos de relações binárias (SMITH *et al.*; 2005):

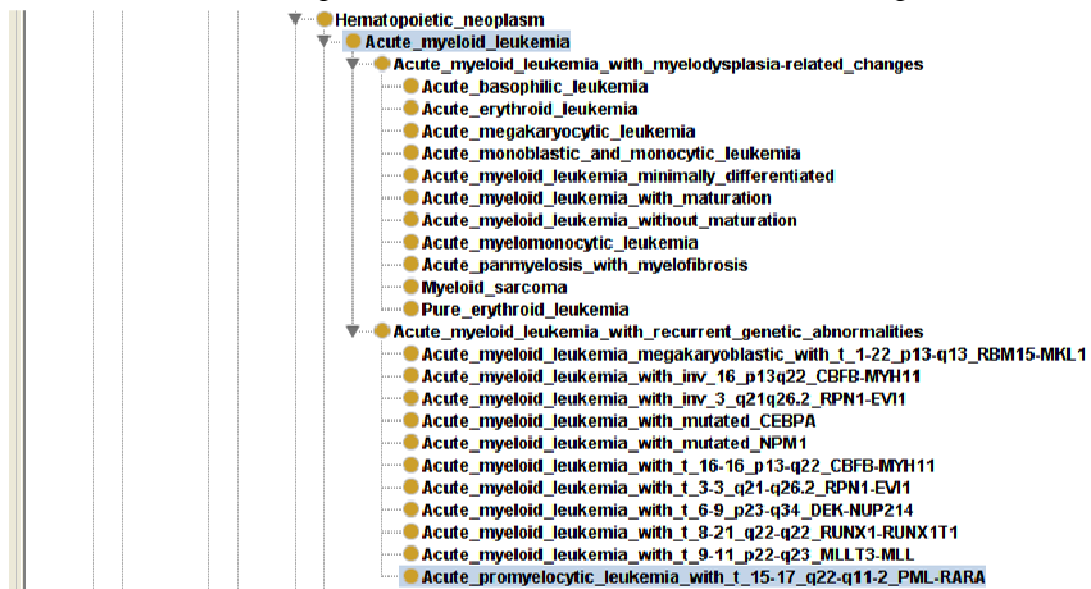
- a) $\langle class, class \rangle$ = de termo para termo, relação **is-a**, relação entre duas classes distintas.
- b) $\langle instance, class \rangle$ = **instance-of**, relação entre um particular e uma classe.
- c) $\langle instance, instance \rangle$ = a relação parte todo: **part-of**, relação entre dois particulares.

Para definir através de relações é necessário definir as entidades. Tudo que existe no espaço temporal do mundo é uma entidade, a qual pode ser um *continuate* (endurante) ou um *ocorrente* (perdurante) (GRENON; SMITH, 2004, p.143). Os continuantes ou endurantes são as entidades que mantêm sua existência através do tempo ao se submeter a diferentes tipos de mudanças, incluindo mudanças de lugares; os ocorrentes ou perdurantes são entidades que se desdobram em fases temporais sucessivas (SMITH *et al.*, 2005).

Na figura 1 apresenta-se a hierarquia da LMA na BLO. Através desta hierarquia, pode-se definir a primeira relação da LMA $\langle class, class \rangle$ através da relação **is-a**, onde *Acute myeloid leukemia is-a hematopoietic neoplasm*. No contexto da BLO, a LMA é uma doença que por ser classificada como um *continuate* (continua a existir), *dependente* (uma vez que depende do organismo para existir), e *disposição* (disposição que qualquer pessoa apresenta em adoecer). Outra relação possível a LMA é a relação $\langle instance, instance \rangle$ **derives-from** (**deriva-de**), *Acute Myeloid Leukemia derives-from hematopoietic stem cell*⁹, uma relação que envolve dois continuantes materiais distintos. A derivação é uma relação entre instâncias, na qual um simples continuante cria uma pluralidade de outros continuantes. Um exemplo citado por Smith *et al.* (2005) são as células proimielocíticas que derivam das mieloblásticas.

⁹ Células-tronco hematopoéticas.

FIGURA 1 - Hierarquia dos termos sobre Leucemias Mieloide Agudas na BLO



Nota: Primeira faixa a definir com 24 classes.

Fonte: Termos da BLO no Protege (ALMEIDA *et al.*,2013).

A formulação de definição para o termo LMA, seguindo os princípios ontológicos é a apresentada a seguir:

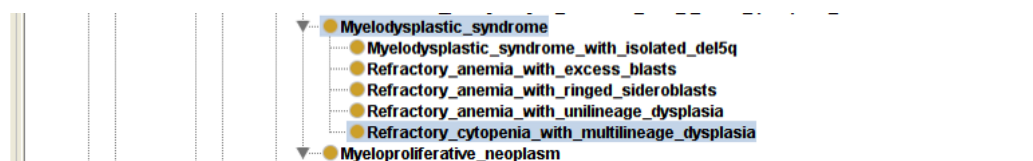
Acute myeloid leukemia *Df* = *A leukemia that occurs when a hematopoietic stem cell undergoes malignant transformation into a primitive, differentiated cell with abnormal longevity and with abnormal proliferation of myeloid cells lineage.*

Por herança, a LMA irá receber as características do termo superior (*hematopoietic*), que será comum aos outros termos da hierarquia das neoplasias do sangue na BLO. A diferença da LMA para outros tipos de leucemias é a célula de linhagem mieloide.

As relações *is-a* na hierarquia da LMA são apresentadas nas figuras 1, 2 e 3:

- Primeira classe da hierarquia: *Acute myeloid leukemia is-a hematopoietic neoplasm*
- Segunda classe da hierarquia: *Acute Myeloid Leukemia with myelodysplasia-related changes is-a Acute Myeloid Leukemia.*
- Terceira classe da hierarquia: *Acute basophilic leukemia is-a Acute Myeloid Leukemia with myelodysplasia-related changes.*

FIGURA 2 - Segunda faixa a definir com 5 classes.



Fonte: Termos da BLO no Protege (ALMEIDA *et al.*,2013).

- d) Primeira classe da hierarquia: *Myelodysplastic syndrome is-a hematopoietic neoplasm.*
- e) Segunda classe da hierarquia: *Myelodysplastic syndrome with isolated del5q is-a Myelodysplastic syndrome.*

FIGURA 3 - Terceira faixa a definir com 11 classes.



Fonte: Termos da BLO no Protege (ALMEIDA *et al.*,2013).

- f) Primeira classe da hierarquia: *Myeloproliferative neoplasm is-a hematopoietic neoplasm.*
- g) Segunda classe da hierarquia: *Atypical chronic Myeloid Leukemia is-a Myeloproliferative neoplasm.*

3 PROCEDIMENTOS METODOLÓGICOS PARA CRIAÇÃO DE DEFINIÇÕES

A pesquisa busca a construção de uma metodologia de criação de definições. No contexto da pesquisa em andamento, os seguintes procedimentos têm sido adotados:

- a) Obtenção da amostra: a amostra terminológica das LMA foi definida pela BLO na qual para a classe das LMA, com um total de 40 termos subdivididos em 3 faixas conforme mostram nas figuras (1,2 e 3).
- b) Definição das formas de aquisição de conhecimento e citação as fontes de informação bibliográficas utilizadas.
- c) Criação de critérios para a formulação das definições em linguagem textual com base na proposta de formulação de definições de Michael, Mejino Junior e Rosse (2001), Köhler *et al.* (2006), Smith *et al.* (2005), dentre outros.

- d) Validação das definições em linguagem textual junto ao especialista oncologista.
- e) Definição de critérios para converter definições de linguagem textual para definições adequadas para ontologias através da linguagem lógica (formal).
- f) Estabelecimento dos relacionamentos necessários além da relação *is-a* para que se possa esclarecer a semântica e assim as características sejam passadas, por herança, aos termos inferiores da hierarquia.
- g) Explicitar e sistematizar os critérios obtidos para formulação de definições para utilização no âmbito da Ciência da Informação.

Os resultados parciais da pesquisa, de acordo com os passos apresentados acima, apontam a possibilidade de obter passos sistemáticos para a criação de definições formais para uso no âmbito da CI. A seguir, apresentam-se uma prévia desses passos desenvolvidos até o momento e ainda em fase de teste no contexto das leucemias:

- a) Separar o termo
- b) Obter uma definição preliminar sobre o significado do termo em algum dicionário
- c) Estabelecer o genus superior no contexto de uso do termo
- d) Estabelecer a característica essencial, distinguindo o genus da espécie
- e) Formular a primeira versão da definição na forma: $S = Def. um G o qual Ds$, onde “G” (para: genus) é o termo pai de “S”; “S” (para: espécies) na ontologia de referência correspondente; e S e G são tipos
- f) Verificar se a definição é uma declaração de condições necessárias e suficientes
- g) Verificar princípio da não circularidade
- h) Verificar herança múltipla
- i) Verificar princípio da substituição
- j) Verificar princípio do desdobramento

4 CONSIDERAÇÕES FINAIS

O presente estudo descreveu pesquisa em andamento visando a obtenção de princípios metodológicos para a criação de definições. Buscou-se definir as LMA em linguagem textual, bem como convertê-las para linguagem de lógica formal no âmbito de projeto denominado de *Blood Ontology* (BLO), um recurso que permite a exploração de informações relevantes para pesquisa científica e manipulação de sangue humano (ALMEIDA *et al.*; 2013). Ao criar

definições sobre os termos dos diversos tipos de LMA foram considerados alguns pressupostos:

1. Como as leucemias são definidas em dicionários, manuais e livros sobre o tema?
2. Que tipo de endurantes e perdurantes são necessários para a classificação de leucemias?
3. Quais relacionamentos são necessários? Que tipos de classes existem?
4. Como estas definições devem ser formuladas?

As metodologias propostas para construção de ontologias ainda não contemplam diretrizes para a criação de definições. Faltam orientações aos cientistas da informação para construção de ontologias de qualidade, onde as definições se configuram como item de importante expressividade semântica. Assim, para construir ontologias em domínios específicos é preciso desenvolver padrões definitórios. Para se alcançar estes padrões é preciso que a definição apresente alguns elementos: gênero próximo, diferença específica, componentes, etapas e finalidade de aplicação, dentre outros. Definições são fundamentais para construção de ontologias por propiciarem a possibilidade de compatibilização semântica ao descrever o conteúdo de um termo. Com isso os agentes inteligentes podem entender o termo e estabelecer inferências sobre o significado (CAMPOS, 2010).

As relações nas ontologias são importantes para formulação de definições formais. Realizar definições em ontologias biomédicas é uma tarefa árdua, complexa e que consome tempo. Definir clinicamente o tipo de leucemia é uma tarefa importante para se estabelecer o tratamento do paciente. Assim as formulações de definições em ontologias biomédicas ultrapassam as fronteiras da organização do conhecimento e desempenham um papel importante no diagnóstico e tratamento do câncer.

São necessários mais estudos sobre como formular definições na área de ontologias biomédicas. Existem algumas poucas experiências, como as da FMA e da GO, que representam esforços para corrigir erros de circularidade e intangibilidade. Na continuidade dessa pesquisa, espera-se contribuir com princípios sistemáticos para a criação de definições em ontologias, bem como fornecer subsídios para profissionais atuando com ontologias no âmbito da CI.

REFERÊNCIAS

- ALMEIDA, M.B.; *et al.*A ontologia do sangue: uma iniciativa para representação e organização do conhecimento sobre Hematologia e Hemoterapia. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**. Rio de Janeiro. v.7, n.1, mar., 2013.
- ALMEIDA, M. B.Revisiting Ontologies: a necessary clarification. **Journal of the American Society of Information Science and Technology**. New York. v. 64, n. 8. p. 1682-93. 2013.
- ALMEIDA, M.B.; *et al.***The Blood Ontology**: an ontology in the domain of hematology.2011. Disponível em : <http://ceur-ws.org/Vol-833/>. Acesso em abr. 2014.
- ARBER, D.A.; COUSAR, J.B. Hematopoietic Tumors: Principles of Pathologic Diagnosis. Chapter 71. In: GREER, J.P.; *et al.* **Wintrobe's Clinical Hematology**. 13th. Philadelphia: Lippincott Williams & Wilkins, 2013.
- CAMPOS, M.L.A.A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 6., 2005, Florianópolis, SC.. **Anais... GT 2: Organização do Conhecimento e Representação da Informação**.
- CAMPOS, M.L.A.O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, Belo Horizonte.v.15, n.1, p.220-238. abr.2010.
- DAHLBERG, I. A Referent-oriented analytical concept theory of interconcept. **International Classification**. v.5, n.3, p.142-150, 1978a.
- DAHLBERG, I.Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro. v.7,n.2, p. 101-107, 1978b.
- GRENON, P.; SMITH, B. SNAP and SPAN: Towards Dynamic Spatial Ontology **Spatial Cognition and Computation**. v. 4, n.1, p.131-177.2004. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.71.9297>. Acesso em: jul 2014.
- GUPTA, A. Definitions. In: ZALTA, E.N. (ed.).**The Stanford Encyclopedia of Philosophy**. UR. Stanford:The Metaphysics Research Lab, 2008. Disponível em: <http://plato.stanford.edu/entries/definitions/> .Acesso em: 22 abr. 2014.
- HJØRLAND, B. Concept Theory. **Journal Of The American Society For Information Science And Technology**. v.60, n.8, p.1519–1536, 2009.
- KAMPEN, K.R. The Discovery and early understanding of leucemia.**Leukemia Research**. v.36, n.1, p.6-13, Jan.2012.
- KÖHLER, J.; *et al.*Quality control for terms and definitions in ontologies and taxonomies. **BMC Bioinformatics**. v.7, p. 212. Apr. 2006.
- MAYO CLINIC. **Acute Myeloid Leukemia**. Disponível em : <http://www.mayoclinic.org/diseases-conditions/acute-myelogenous-leukemia/basics/definition/con-20043431>. Acesso em: fev.2014.

MEDICAL SUBJECT HEADINGS(MeSH). **Mesh Definitions:** Acute Myeloid Leukemia. Disponível em: <http://www.ncbi.nlm.nih.gov/mesh>. Acesso em abr. 2014.

MICHAEL, J; MEJINO JL JR.; ROSSE, C. The role of definitions in biomedical concept representation. Proceedings / American Medical Informatics Association ... Annual Symposium. **AMIA Symposium**. p. 463-7.2001.

NAJFELD, V. Conventional and molecular cytogenetic basis of hematology malignancies. In: HOFFMAN, R.; *et al.* (ed.). **Hematology:** basic principle and practice. 5th.ed. Churchill Livingstone: Elsevier,2009.cap. 55, p.791-838.

NATIONAL CANCER INSTITUTE (NCI). **Acute Myeloid Leukemia**.2014. Disponível em: <http://www.cancer.gov/cancertopics/pdq/treatment/adultAML/healthprofessional/page1>. Acesso em: fev. 2014.

NATIONAL CANCER INSTITUTE (NCI).**What You Need To Know About Leukemia**. NIH Publication No. 13-3775 Revised September 2013, Digital Edition. Disponível em : <http://www.cancer.gov/cancertopics/wyntk/leukemia.pdf>. Acesso em abr. 2014.

NORMA ISO 704: 2000.**Terminology work:**Principles and methods. 2.ed. Genève: International Standard Organization, 2000.

ONTOBEE. **Leukemia**. Disponível em: http://www.ontobee.org/browser/rdf.php?o=DOID&iri=http://purl.obolibrary.org/obo/DOID_1240. Acesso em: maio 2014a.

ONTOBEE.**Welcome to Ontobee!** Disponível em: <http://www.ontobee.org/>. Acesso em: maio 2014b.

SMITH, B.; *et al.*Relations in biomedical ontologies. **Genome biology** . v.6, n.5, p.R46.2005.

SMITH, B.; *et al.*The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature biotechnology**. v.25, n.11, p.1251-5.nov. 2007.

SMITH, B.Introduction to the Logic of Definitions. In: INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIE, DO 2013, July 7, Montreal, **Anais...** 2013.Disponível em : http://ceur-ws.org/Vol-1061/Paper5_DO2013.pdf. Acesso em maio 2014.

SOERGEL, D.; *et al.***Reengineering Thesauri for New Applications:** the AGROVOC Example.p.1-23.Disponível em:<http://www.fao.org/docrep/008/af234e/af234e00.HTM>. Acesso em: set. 2014.

SWARTZ, N. **Definitions, Dictionaries, and Meanings**. This revision: November 8, 2010. Disponível em: <http://www.sfu.ca/~swartz/definitions.htm>.Acesso em: maio 2014.

USCHOLD, M.**Building Ontologies:** Towards a Unified Methodology.1996. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.9075> . Acesso em fev. 2013.

VICKERY, B.C. Ontologies. **Journal of Information Science**, v. 23. n.4, p. 227-286, 1997.

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, 2009. cap. 59, p. 921-932.

XIANG, Z.; *et al.* Ontobee: A Linked Data Server and Browser for Ontology Terms. **Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO)**, July 28-30, 2011, Buffalo, NY, USA. p. 279-281. Disponível em: http://www.ontobee.org/Ontobee_ICBO-2011_Proceeding.pdf. Acesso em jul. 2014.

Agradecimentos

Ao Dr. Joaquim Caetano de Aguirre Neto, médico oncologista pediátrico, pelo suporte e apoio na validação das definições em linguagem textual das leucemias.