



ARTIGO DE PESQUISA

Comunicação entre sistemas de informação médicos
princípios para elaborar definições

Amanda Damasceno de Souza¹  <https://orcid.org/0000-0001-6859-4333>

Maurício Barcellos Almeida²  <https://orcid.org/0000-0002-6153-1011>

¹ Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil. / e-mail: amanda@ufmg.br

² Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil. / e-mail: mba@eci.ufmg.br

RESUMO

A formulação de definições de termos médicos para uso em sistemas de informação baseados em ontologias é fundamental para a continuidade do cuidado quando um paciente se move entre diferentes unidades médicas. Entretanto, a prática de criar boas definições não é uma tarefa trivial e a literatura não contempla metodologias para tal. Este artigo tem como objetivo apresentar um método que estabelece princípios para sistematização do processo de criação de definições padronizadas, as quais possam proporcionar uma melhor comunicação entre sistemas de informação médicos, e, portanto, melhor atendimento ao cidadão. A criação e teste do método foi conduzido no domínio das leucemias, no âmbito do projeto *Blood Ontology*. Os resultados mostram que definições de entidades biomédicas são definidas por uma variedade de critérios, que nem sempre atendem aos requisitos de representação ontológica para a adequada comunicação entre sistemas médicos. Problemas verificados dizem respeito a identificação de características essenciais de cada entidade e envolvem a circularidade, a intangibilidade e a complexidade em diagnosticar doenças complexas como leucemias. Espera-se que essa experiência possa contribuir para melhorias nos projetos em que a Ciência da Informação apoia a medicina e os cuidados à saúde na busca por melhor atendimento ao cidadão.

PALAVRAS-CHAVE

Sistemas de organização do conhecimento. Ontologias. Conceitos. Medicina.

Communication among medical information systems
principles for elaborating definitions

ABSTRACT

The formulation of definitions for medical terms to be applied in ontology-based information systems is a fundamental activity for supporting continuing care, in cases in which a patient moves among different medical units. However, the practice of creating good definitions is not a trivial task, and the literature does not provide clear methodologies. This article aims to present a method to establish principles for systematizing the process of creating standardized definitions, which can be capable to provide the expected communication between the medical information systems and, therefore, to foster better services for citizens. The creation and testing of such method was conducted in the field of leukemias in the scope of the *Blood Ontology* project. The results show that definitions of biomedical entities are defined in a variety of criteria, which do not always meet the ontological representation requirements for adequate communication between medical information systems. Some well-known issues already identified are the identification of essential features of an entity, in addition to involve circularity, intangibility as well as the complexity in diagnosing manifold diseases such leukemias. We

hope this experience can contribute to improvements in research projects in which Information Science supports Medicine and Healthcare with the aim of providing a better service to the citizen.

KEYWORDS

Knowledge organization systems. Ontologies. Concepts. Medical Sciences.



JITA: ID. Knowledge representation.

1 INTRODUÇÃO

A comunicação entre unidades de saúde é essencial para os cuidados continuados a saúde, como por exemplo, em longos tratamentos de câncer. Essa comunicação atualmente é feita via sistemas de informação, os quais normalmente não se conectam adequadamente por diversos fatores, dentre os quais as diferenças conceituais entre termos e suas definições. No contexto dos estudos sobre ontologias, buscam-se criar representações bem fundamentadas do domínio médico de forma a permitir comunicação em diferentes espaços geográficos e temporais.

As ontologias no domínio biomédico são importantes para conectar elementos comuns entre dados da biologia e área afins. Para isso, as ontologias biomédicas são estruturadas logicamente de forma hierárquica, conectando classes que buscam representar as entidades e seus relacionamentos. Neste contexto, foi criado o *Open Biological and Biomedical Ontologies (OBO) Foundry* (2018) com o objetivo de prover princípios para ajudar na padronização de ontologias no domínio biomédico (MEEHAN *et al.*, 2011). Segundo Seppälä e Ruttenberg (2013), as ontologias baseadas nos princípios da *OBO Foundry* incluem tanto definições textuais (linguagem natural) como definições formais (lógicas). Definições são importantes para padronizar e representar dados, vocabulários e terminologias, assim como para criar taxonomias, com objetivo de promover a comunicação entre sistemas (KEIZER; ABU-HANNA; ZWETSLOOT-SCHONK, 2000).

Especificamente em ontologias que fundamentam os modernos sistemas de informação médicos, as definições dos termos desempenham uma importante função, servindo a vários propósitos: permitem criar hierarquias, verificar a validade dos arranjos das classes e categorias, compreender o significado de termos especializados e integrar dados (IWOOD, 2014). Na Ciência da Informação (CI) a teoria clássica sobre definições foi preconizada pela alemã Ingetraut Dahlberg por meio da **Teoria do Conceito** como uma questão importante na construção dos Tesouros (DAHLBERG, 1978a,b). Para Dahlberg (1978a, p.6) as definições apresentam uma função importante na formulação dos conceitos, pois “são pressupostos indispensáveis na argumentação e nas comunicações verbais e que constituem elementos necessários na construção de sistemas científicos”.

Entretanto, a questão que se coloca é como formular essas definições de forma alinhada com os modernos sistemas de informação (KÖHLER *et al.*, 2006). A dificuldade em se criar definições em ontologias se relaciona ao fato de que, além complexidade e de custo da tarefa, ela é passível de erros e precisa ser realizada por pessoas treinadas (TSATSARONIS *et al.*, 2013). A ontologia apresenta as características necessárias para resolver problemas referentes à definição dos termos (SPEAR, 2006). Entretanto, formular definições exige um conhecimento específico de suas teorias e requisitos, tanto do ponto de vista da medicina quanto da Ciência da Informação.

Para redigir definições com critérios bem formados, é preciso começar pela definição em linguagem natural, conforme sugere Gruber (1993). Em seguida, será necessário formular definições em linguagem de representação lógica, as quais são as definições utilizadas pelos sistemas de informação computacionais da era da Web Semântica. As metodologias para construção de ontologias como artefatos sugerem a presença de definições, sem explicar, entretanto, como devem ser formuladas. Esse artigo é uma iniciativa nesse sentido, no âmbito da Ciência da Informação. A padronização de definir é essencial para que a definição seja rigorosa e contenha apenas o texto necessário, o que permite que ela seja usada para localizar o termo na estrutura taxonômica da ontologia.

O restante do presente artigo foi organizado da seguinte forma: a seção 2 traz uma breve introdução sobre definições e suas características; a seção 3 introduz o domínio das leucemias; a seção 4 apresenta o método sistematizado aplicado a termos do domínio das leucemias; a seção 5 discute os resultados e a seção 6 tece breves considerações finais. Cabe destacar que, como a pesquisa é parte de um projeto internacional, algumas definições e termos são deixados no seu original em língua estrangeira sem prejuízo para o entendimento médico.

2 CARACTERÍSTICAS DAS DEFINIÇÕES

Crítérios relevantes para o projeto de elaboração de ontologias são a clareza e a coerência: “todas as definições devem ser documentadas em linguagem natural (GRUBER, 1993, p.2 e 3). Também Uschold (1996, p.12 e 13) apresenta vários critérios básicos para a criação de definições, dentre eles clareza e coerência, como citados a seguir:

- a) Clareza: definições devem ser claras e inequívocas quando expressas em linguagem natural ou formalmente codificadas. Recomenda-se utilizar exemplos para ilustrar o que se pretende e também exemplos negativos para deixar claro o que não se pretende;
- b) Consistência e coerência: uma ontologia deve ser internamente e externamente consistente, e assim a circularidade deve ser evitada; deve evitar-se também introduzir novos termos sem que estes estejam esclarecidos;
- c) Extensibilidade e reutilização: ontologia deve ser projetada de forma que permita a sua reutilização e extensibilidade, e para tal, definições devem ser bem formadas.

Mesmo com a existência de tais recomendações na literatura, diversos problemas persistem quando se definem termos de um domínio complexo com a medicina. Seppälä e Ruttenberg (2013) realizaram um estudo sobre práticas para criação de definições em ontologias, nos quais relatam dimensões principais de problemas relacionados a definições (Quadro 1).

Quadro 1. Problemas de definições

Conteúdo	Lógica	Escrita e estilo	Cobertura
Informação insuficiente, muito informativo ou muito complexo, desatualizado ou obsoleto, ausência de padrão	Imprecisa ou vaga, circular, contraditória.	Mal redigida, estilo inconsistente.	Várias definições similares ausência de definições

Fonte: adaptado de Seppala e Rutemberg (2013).

No quadro 2 são exemplificados problemas de definições da *Gene Ontology* (GO)¹ e do *NCI-Thesaurus*²:

¹ <http://geneontology.org/>

² <https://ncithesaurus.nci.nih.gov/ncitbrowser/>

Quadro 2. Exemplos de Problemas de definições

Conteúdo	Lógica	Cobertura
<p>GO:0050566 termo: <i>asparaginyl-tRNA synthase (glutamine-hydrolyzing) activity</i></p> <p>definição: <i>Catalysis</i> Cyc:6.3.5.6-RXN</p>	<p>GO:0042270 termo: <i>Protection from natural killer cell mediated cytotoxicity</i></p> <p>definição: <i>The process of protecting a cell from cytotoxicity by natural killer cells</i></p>	<p>NCI Thesaurus termo: Acute Myeloid Leukemia:</p> <p>definição: <i>A clonal expansion of myeloid blasts in the bone marrow, blood or other tissues. The classification of acute myeloid leukemias (AMLs) encompasses four major categories: 1) AML with recurrent genetic abnormalities; 2) AML with multilineage dysplasia; 3) Therapy-related AML; 4) AML not otherwise specified. The required bone marrow or peripheral blood blast percentage for the diagnosis of AML is 20% (WHO classification).</i></p> <p>Outras definições: <i>An aggressive (fast-growing) disease in which too many myeloblasts (immature white blood cells that are not lymphoblasts) are found in the bone marrow and blood.</i></p> <p><i>A rapidly progressive cancer of the blood and bone marrow consisting of the proliferation of abnormal myeloblasts, which are immature, dysfunctional white blood cells.</i></p>
<p>GO:0001655 termo: urogenital system development</p> <p>definição: <i>the development of the urogenital system</i></p>	<p>GO:0006190 termo: inosine salvage</p> <p>definição: <i>Any process that generates inosine, hypoxanthine riboside, from derivatives of it without de novo synthesis.</i></p>	<p>NCI Thesaurus termo: Chronic Myelogenous Leukemia</p> <p>definição: <i>A chronic myeloproliferative neoplasm characterized by the expression of the BCR-ABL1 fusion gene. It presents with neutrophilic leukocytosis. It can appear at any age, but it mostly affects middle aged and older individuals. Patients usually</i></p>

		<p><i>present with fatigue, weight loss, anemia, night sweats, and splenomegaly. If untreated, it follows a biphasic or triphasic natural course; an initial indolent chronic phase which is followed by an accelerated phase, a blast phase, or both. Allogeneic stem cell transplantation and tyrosine kinase inhibitors delay disease progression and prolong overall survival.</i></p> <p><i>Outra definição: A slowly progressing disease in which too many white blood cells are made in the bone marrow.</i></p>
<p>NCI Thesaurus Term: Tuberculosis</p> <p>definição: <i>A chronic, recurrent infection caused by the bacterium Mycobacterium tuberculosis. Tuberculosis (TB) may affect almost any tissue or organ of the body with the lungs being the most common site of infection. The clinical stages of TB are primary or initial infection, latent or dormant infection, and recrudescent or adult-type TB. Ninety to 95% of primary TB infections may go unrecognized. Histopathologically, tissue lesions consist of granulomas which usually undergo central caseation necrosis. Local symptoms of TB vary according to the part affected; acute symptoms include hectic fever, sweats, and emaciation; serious complications include granulomatous erosion of pulmonary bronchi associated with hemoptysis. If untreated, progressive TB may be associated with a high degree of mortality. This infection is frequently observed in immunocompromised individuals with AIDS or a history of illicit IV drug use.</i></p>	<p>Term: Tuberculosis</p> <p>definição: <i>A chronic, recurrent infection caused by the bacterium Mycobacterium tuberculosis.</i></p>	<p>NCI Thesaurus Termo: Acute Leukemia</p> <p>definição: <i>A clonal (malignant) hematopoietic disorder with an acute onset, affecting the bone marrow and the peripheral blood. The malignant cells show minimal differentiation and are called blasts, either myeloid blasts (myeloblasts) or lymphoid blasts (lymphoblasts).</i></p> <p>Outra definição: <i>A rapidly progressing cancer that starts in blood-forming tissue such as the bone marrow, and causes large numbers of white blood cells to be produced and enter the blood stream.</i></p>

Fonte: Köhler et al. (2006, p.9), National Cancer Institute Thesaurus (2020) e Smith (2013).

Spear (2006) explica que, para sistemas de informação distintos que armazenam informação biomédica sejam compatíveis ou interoperáveis, é necessário que se utilizem terminologia consensual e definições consistentes. Para isso, os termos precisam ser internamente coerentes, respaldados e alinhados com fatos da realidade das ciências biomédicas. Dessa forma, princípios básicos precisam ser adotados para a formulação de definições adequadas, como os citados no restante da presente seção.

2.1 Princípio das condições necessárias e suficientes

Definições criadas de acordo com princípios aristotélicos enfatizam condições necessárias e suficientes. Nesse caso, uma definição é de fato uma declaração de condições necessárias e suficientes (SMITH, 2013). Por exemplo: “ser uma entidade A é uma condição necessária para ser uma entidade, ou seja, todo B é um A; ser uma entidade A é uma condição suficiente para ser uma entidade B, ou seja, todo A é um B; [...] se essas duas declarações se mantêm, definir um A é definir algo que satisfaz um B [...]. Isto porque A é um termo mais difícil de ser compreendido do que B” (SMITH, 2013, p.1).

Para Swartz (2010), a relação entre “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. Nessa afirmação, a implicação do inverso de uma afirmação é o resultado do invertido de suas duas partes.

2.2 Princípio da herança única

O papel de definições em uma ontologia é especificar como especificar as características dos termos representativos de entidades de forma consistente, ou seja, de forma que se assegure a herança transitiva pela hierarquia. A consistência nas definições necessita de um ponto de vista unificador – o contexto – na representação, o qual deve manter-se em toda a ontologia. Ao obedecer a esse requisito, a posição de um conceito na hierarquia será determinada pela sua própria definição e pela definição do termo superior.

No princípio de herança, dois aspectos das definições aristotélicas são importantes: o *Genus* e a *Differentia*. O *Genus* é a característica essencial que todos os subterramos da hierarquia irão assumir, ou seja, uma definição existente contribui para a parte de uma nova definição; a *Differentia* busca definir as características essenciais que distinguem um tipo de entidade de outros tipos, ou seja, atribui uma característica única ao termo (MICHAEL; MEJINO JUNIOR; ROSSE, 2001, p.463).

Segundo Seppälä, Schreiber e Ruttenberg (2014), a primeira parte da definição é o *Genus* e a segunda parte é a *Differentia*. Definições devem ser formuladas seguindo a estrutura de *genus + differentia* (SEPPÄLÄ, 2016). A herança múltipla acontece quando uma entidade tem dois pais na hierarquia. Ao evitar a herança múltipla, evitam-se também ambiguidades e problemas de entendimento.

2.3 Princípio da não-circularidade

Köhler et al. (2006, p.2) recomendam diversas regras baseadas também nos princípios aristotélicos para formulação de definições em ontologias, dentre elas evitar circularidade. A circularidade ocorre quando um termo é usado para definir ele mesmo, por exemplo, telefone sem fio é um telefone que não tem fio.

Atenção assim é devida a regra da circularidade para evitar que uma definição utilize seus próprios termos para definir a si mesma. Na ontologia, é importante seguir princípios de não circularidade, possibilitando que as definições sejam escritas de forma clara para que se compreenda o real significado do termo. A não-circularidade enseja definições concisas, em frases curtas e completas (KÖHLER et al., 2006).

2.4 Princípio da intangibilidade

Ao formular definições em ontologias, é preciso evitar a assim chamada “linguagem figurativa ou obscura”, ou seja, uma definição deve esclarecer o significado de uma palavra desconhecida, não dificultar o entendimento. Essa regra é denominada de princípio da intangibilidade, que sugere que a definição deve usar termos inteligíveis, ou seja, que sejam facilmente compreendidos, uma vez que uma definição deve ser compreendida por pessoas que não sejam especialistas do domínio (KÖHLER et al., 2006).

O princípio da intangibilidade refere-se à qualidade da terminologia, de forma que cada termo definido deve atender aos padrões básicos de compreensão, para que a definição seja inteligível e totalmente compreensível sem que sejam necessárias leituras prévias ou consultas adicionais a fontes de informação. Isso inclui também evitar o uso de jargão técnica (KÖHLER et al. 2006).

3 DEFINIÇÕES NO DOMÍNIO DAS LEUCEMIAS

Para propor um método para formulação de definições em linguagem natural, apropriadas a ontologias de sistemas de informação biomédicos, estudou-se o caso das Leucemias Mieloides Agudas (LMAs) no âmbito de projeto *Blood Ontology* (BLO) (ALMEIDA et al., 2013). A BLO é um vocabulário formal abrangendo o conhecimento em hematologia e hemoterapia, que objetiva reunir, organizar e facilitar a manipulação de dados sobre sangue humano. A ontologia descreve neoplasias hematológicas que compreendem leucemias e linfomas. Na presente pesquisa, o objeto de estudo será a definição dos termos das leucemias presentes na BLO.

Ao definir um domínio, uma etapa importante é aquisição de conhecimento, que possibilita conhecer o domínio em questão. Essa etapa permite entender que leucemias são doenças complexas e, em função disso, nos últimos 30 ou 40 anos, a análise citogenética das neoplasias hematológicas foi uma das áreas da hematologia que mais se desenvolveram (NAJFELD, 2009). Leucemias são classificadas de acordo com a rapidez da diferenciação celular em: i) agudas, em que células anormais se caracterizam pelo desenvolvimento rápido; ii) crônicas, em que células anormais se desenvolvem mais lentamente. Leucemias são ainda diferenciadas pelo tipo de glóbulo branco afetado em: iii) mielóide, que se inicia nas células mieloides, também chamadas de mieloblásticas; iv) linfóide, que se inicia nas células linfóides e é conhecida como linfóide, linfoblástica ou leucemia linfocítica (NATIONAL CANCER INSTITUTE, NCI, 2013b).

A LMA refere-se assim a um grupo de doenças heterogêneas com respeito à clonalidade, alterações cromossômicas e resposta ao tratamento. É uma doença que exibe uma heterogeneidade fenotípica, ou seja, que se refere às reais propriedades observadas em um organismo, tais como, morfologia, desenvolvimento ou comportamento; e genotípica, ou seja, relativa à genética hereditária do indivíduo. É caracterizada pela proliferação e acúmulo de células hematopoiéticas imaturas da medula óssea e do sangue. Essas células malignas

substituem gradualmente os precursores de eritróide normais, mielóides e megacariócitos, além de inibir o crescimento e maturação. A sua definição das LMA é feita com base no diagnóstico estabelecido por critérios de morfologia, imunofenotípica, citogenética e molecular (ARBER; COUSAR, 2013). As células blásticas normalmente compõem de 1% a 5% de células de medula. Quando o número de blastos se encontra acima de 20% (critério definido pela Organização Mundial de Saúde), o diagnóstico é LMA. Esse conhecimento médico é necessário ao profissional de Ciência da Informação de forma que ele possa elaborar definições no domínio das leucemias.

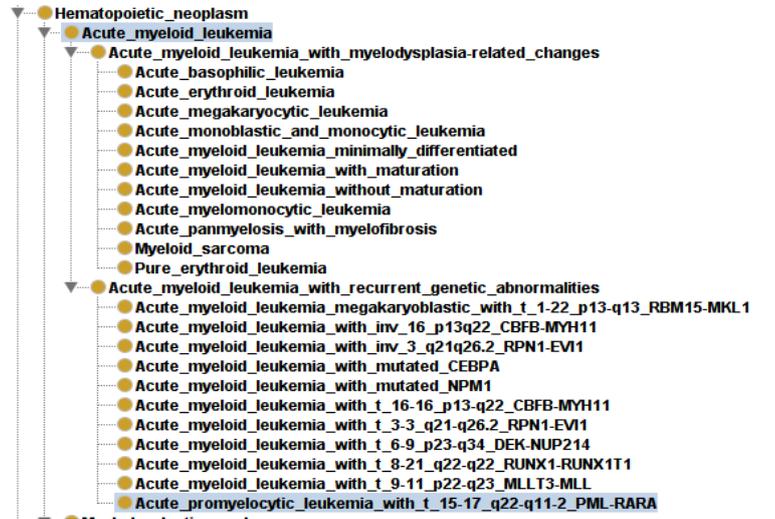
4 SISTEMATIZAÇÃO DE DEFINIÇÕES EM ONTOLOGIAS BIOMÉDICA

As etapas do método de formulação de definições no domínio das LMAs, são apresentadas no restante da presente seção.

4.1 Amostra para pesquisa

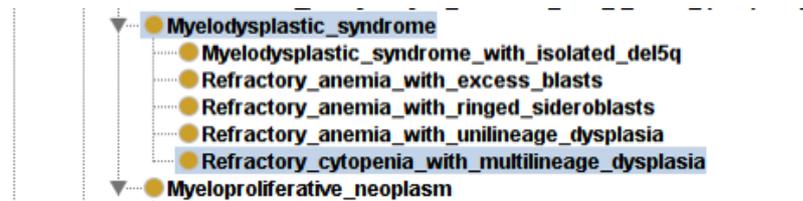
O recorte da amostra terminológica das LMAs foi definido pelo conjunto de termos que compreendia as neoplasias hematológicas da BLO, restrito a análise das leucemias que apresentavam como progenitor a linhagem mieloide, divididas em três classes: i) para a classe das LMA, há um total de 25 termos; ii) para a síndrome mielodisplásica, há 6 termos, e iii) para as neoplasias mieloproliferativas há 12 termos. O total a definir soma 43 termos, conforme mostram as figuras 1, 2 e 3, respectivamente. Trata-se de uma amostra por conveniência e o critério para extraí-la foi favorecer termos do domínio do câncer da BLO, de forma que o total de 43 termos abrangem as neoplasias do sangue. O enfoque principal da pesquisa foi a primeira classe, as LMAs, mas na classe das neoplasias mieloproliferativas também foram definidas as leucemias crônicas.

Figura 1. Fragmento da primeira faixa a definir com 25 classes



Fonte: Fragmento da BLO no Protégé – Almeida et al., (2013).

Figura 2. Fragmento da segunda faixa a definir com 6 classes



Fonte: Fragmento da BLO no Protégé – Almeida et al., (2013).

Figura 3. Fragmento da terceira faixa a definir com 12 classes.



Fonte: Fragmento da BLO no Protégé – Almeida et al., (2013).

No item 4.4 deste artigo apresenta-se um exemplo prático de como formular definição, para o termo da amostra: *Acute Myeloid Leukemia*.

4.2 Aquisição de conhecimento para definições

Para aquisição de conhecimento é importante recuperar na literatura, textos e definições do domínio em questão a partir de fontes especializadas. Para o método proposto foram pesquisadas definições das leucemias citadas na amostra da pesquisa nas seguintes fontes

online: *NCI Thesaurus* (NCIt); *NCI Dictionary of Cancer Terms*; *Medical Subject Headings* (MeSH); *Medscape*, *Ontobee*; *Disease Ontology* (DO); e na fonte textual *Patologia de Foucar et al.*, (2012).

Neste estudo foram pesquisadas fontes de informação em saúde que apresentavam como tema, o câncer, obras indicadas pelos especialistas da área de oncologia, além de material utilizado na BLO. Outra etapa importante na aquisição de conhecimento é reutilizar conhecimento de outras ontologias, como a *Disease Ontology* e a *Gene Ontology* (XIANG et al., 2011; SCHRIML et al., 2012).

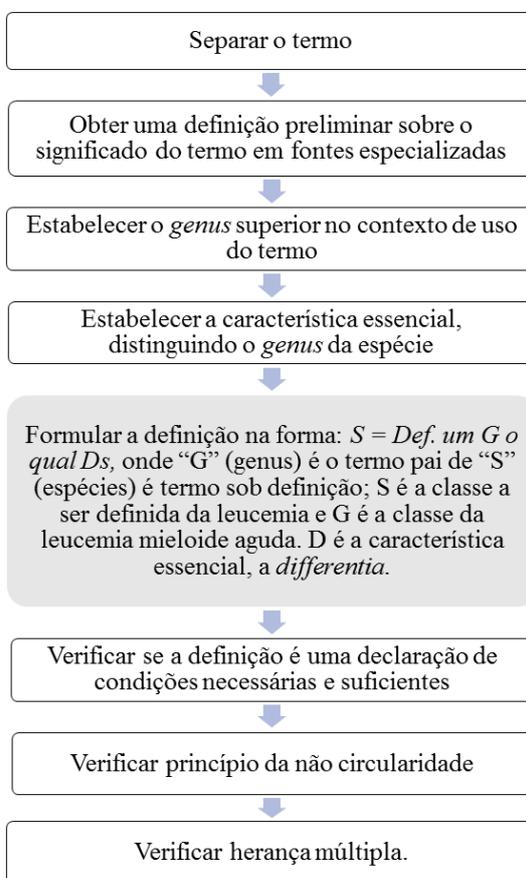
É necessário avaliar as definições existentes na literatura e assim verificar se tais apresentam alguns dos problemas já mencionados, descritos por Seppälä e Ruttenberg (2013).

4.3 Lista de princípios básicos para criar definições em linguagem natural

Ao formular definições em linguagem natural, é preciso obedecer aos princípios de não circularidade e tangibilidade já mencionados e sugeridos por Köhler et al. (2006). Para prover definições que sejam úteis para a compreensão, adota-se linguagem clara e precisa, evitando o uso de termos inerentes ao próprio termo a ser definido, e evita-se linguagem figurativa ou obscura. Ao usar termos técnicos na área biomédica, busca-se descrevê-lo da forma mais acessível possível.

Para a sistematização do método da pesquisa, as etapas foram baseadas em boas práticas identificadas na literatura propostas por pesquisadores diversos como Michael, Mejino Junior e Rosse (2001); Köhler et al. (2006); Smith et al. (2005); Soergel (2014); Seppälä e Ruttenberg (2013); Petrova et al. (2015); Tsatsaronis et al. (2013); Seppälä, Schreiber e Ruttenberg (2014); dentre outros. As etapas do método são apresentadas na Figura 4:

Figura 4. Método de elaboração de definições



Fonte: Dados da pesquisa.

Após a sistematização dos princípios básicos para criar definições em linguagem natural, foi necessário aplicar esses princípios em cada termo da amostra da pesquisa. Apresenta-se a seguir, o resultado do conhecimento adquirido na realização desta etapa da pesquisa.

4.4 Como aplicar o método de definições

A seguir, descreve-se em mais detalhes a forma de aplicação de cada passo da lista de princípios básicos para criar definições textuais (Quadro 3):

Quadro 3. Lista de princípios básicos para criar definições textuais

1) Separar o termo	Listar o termo a ser definido. Na área médica, podem haver termos diferentes ou sinônimos para uma doença.
2) Obter uma definição preliminar	Usar fontes de informação da área específica do domínio: livros, artigos, dicionários, enciclopédias, tesouros, ontologias.
3) Estabelecer o genus superior no contexto de uso do termo	O <i>genus</i> , a característica comum aos termos da hierarquia, definido pela relação <is-a>, por ex.: <i>acute myeloid leukemia is-a hematopoietic neoplasm</i> .

4) Estabelecer a característica essencial, distinguindo o genus da espécie	Utilizar a <i>differentia</i> para definir a característica única e essencial (ou grupo) que distingue um termo do outro. Essa característica essencial, em geral, demanda um especialista.
5) Formular a primeira versão da definição na forma	$S = um\ G\ o\ qual\ Ds$, onde “G” (para: <i>genus</i>) é o termo pai de “S”; “S” (para: espécies) onde S é a classe da leucemia a ser definida, G é a classe mais geral e D é a <i>differentia</i> que distingue a instância de S da instância de D:
6) Verificar condições necessárias e suficientes	Utilizar a expressão: ser um A é condição necessária para ser um B, então cada B é um A. Ser um A é condição suficiente para ser um B, então cada A é um B; isso significa que “A” representa a característica essencial e “B” representa o termo definido.
7) Verificar princípio da não-circularidade	Verificar se a definição utilizou o(s) próprio(s) termo(s) para definir a si mesmo.
8) Verificar presença de herança múltipla	Verificar a presença de nós filhos da hierarquia que possuem dois nós pais.

Fonte: Dados da Pesquisa.

A seguir apresenta-se um exemplo prático real de aplicação do método de elaboração de definições no domínio das leucemias (Quadro 4).

Quadro 4. Elaboração das definições das Leucemias

Etapa	Exemplo	Comentário
1) Separar o termo	<i>Termo: Acute Myeloid Leukemia</i>	Termo da leucemia a ser definido, de acordo com a BLO
2) Obter uma definição preliminar sobre o significado do termo em fontes especializadas	NCI Thesaurus (NCIT): <i>a clonal expansion of myeloid blasts in the bone marrow, blood or other tissues. The classification of acute myeloid leukemias (AMLs) encompasses four major categories: 1) AML with recurrent genetic abnormalities; 2) AML with multilineage dysplasia; 3) Therapy-related AML; 4) AML not otherwise categorized. The required bone marrow or peripheral blood blast percentage for the diagnosis of AML is 20% (WHO classification).</i>	As fontes de informação selecionadas foram: <i>Disease Ontology, Ontobee, MeSH, Medscape, Livro Foucar et al. (2012), NCI-Dictionary of Cancer Terms, NCI-Thesaurus.</i>
3) Estabelecer o genus superior no contexto de uso do termo	<i>Acute myeloid leukemia is a hematopoietic neoplasm.</i>	No caso das leucemias, o <i>genus</i> comum aos termos é uma leucemia mieloide aguda que apresentam como característica uma derivação anormal na linhagem mieloide. Utilizou-se a relação <is-a> para estabelecer o <i>genus</i> .

<p>4) Estabelecer a característica essencial, distinguindo o genus da espécie</p>	<p><i>Derives from an uncontrolled proliferation of the myeloid lineage and their precursors.</i></p>	<p>O fato de simplesmente ser da linhagem mieloide não caracteriza a LMA, mas sim a alteração na linhagem mieloide, pois também existe a diferenciação para linhagem linfóide. No caso das LMAs, a <i>differentia</i> entre as leucemias foi definida pelo tipo de célula envolvida (diferenciação da linhagem mieloide) e porcentagem de células imaturas (blastos).</p>
<p>5) Formular a primeira versão da definição na forma: <i>S = Def. um G o qual Ds</i>, onde “G” (para: <i>genus</i>) é o termo pai de “S”; “S” (para: espécies) onde S é a classe a ser definida, G é a classe mais geral e D é a <i>differentia</i> que caracteriza a instância de S da instância de D.</p>	<p><i>An Acute myeloid leukemia is-a hematopoietic neoplasm that derives from an uncontrolled proliferation of the myeloid lineage and their precursors.</i></p>	<p>Evitar o uso de palavras incomuns, por exemplo, na definição de <i>Myeloid neoplasms with abnormalities of PDGFRA PDGFRB or FGFR1</i>, modificou-se <i>aberrant</i> por <i>abnormal</i>.</p>
<p>6) Verificar condições necessárias e suficientes</p>	<p>Ser uma LMA é condição suficiente para “derivar de uma proliferação sem controle de uma linhagem mieloide e seus precursores”, ou seja, cada LMA deriva de uma proliferação sem controle de uma linhagem mieloide e seus precursores.</p>	<p>Realizar a verificação de herança única para confirmar se a característica definida no item 4, é necessária e suficiente para cada tipo de leucemia e se existem outras leucemias ou outras doenças com essas mesmas características.</p>
<p>7) Verificar princípio da não circularidade</p>	<p>Não parece haver circularidade, mantemos a definição como em (5).</p>	<p>Para corrigir a circularidade de termos mais complexos, as vezes é necessário descrever significados das siglas por extenso e esclarecer a forma de mutação: se era translocação, deleção ou inversão. O termo mais difícil de evitar a circularidade foi <i>Acute myeloid leukemia with mutated CEBPA</i>.</p>

8) Verificar herança múltipla	Herança única: linhagem mieloide.	Por se tratar de doença clonal, poderia ter linhagem mieloide ou linfoide. Assim, foi definida a herança mieloide de acordo com os critérios de classificação da FAB/WHO (2008). Entretanto, Silva <i>et al.</i> (2006) relata a existência de tipos de leucemia aguda com ambas as linhagens (mieloide e linfoide), sendo consideradas leucemias de linhagens mista, híbrida ou bifenotípica.
-------------------------------	-----------------------------------	--

Fonte: Dados da pesquisa.

No item a seguir 4.5, apresenta-se a validação que foi realizada com especialista de domínio do câncer na formulação das definições da amostra da pesquisa.

4.5 Validação das definições com especialista

A validação do conhecimento inserido nas definições foi realizada com base nas condições necessárias e suficientes, com o auxílio de um especialista de formação na área médica, especialista em oncologista. Esse processo já foi previamente testado no projeto BLO e consiste em mostrar possibilidades de definição, e depois solicitar validação e aprovação de especialistas (COELHO, 2012).

Para a validação da definição questionou-se ao médico (especialista de domínio):

- i) é correto afirmar que uma entidade qualquer derivando de uma proliferação sem controle de uma linhagem mieloide e seus precursores é a característica essencial de uma Leucemia Mieloide Aguda?; ii) existem outras entidades que derivam de uma proliferação sem controle e não são Leucemia Mieloides Agudas?

Neste contexto o médico oncologista ressaltou:

Para definir uma LMA, é necessário observar três características: morfológica, citogenética e de imunofenótipo. A LMA é de origem única, a linhagem mieloide, pois se trata de uma doença clonal e assim descende de apenas uma linhagem (a mieloide). A leucemia, quando é bifenotípica ou bi-linhagem, tem duas origens por apresentar várias populações de células clonais... entretanto, este tipo apresenta somente a linhagem mieloide. Mesmo que seja mínima a descendência da linhagem mieloide sua presença sugere a diferenciação.

Com isso conclui-se que a essência da definição está correta e decidiu-se manter a definição tal como foi formulada pelo método, à saber: *an Acute myeloid leukemia is a hematopoietic neoplasm that derives from an uncontrolled proliferation of the myeloid lineage and their precursors*. Processo similar foi repetido para as demais entidades que se separou para fins de definição.

4 DISCUSSÃO

Para elaborar definições, é necessário em primeiro lugar, buscar fontes de conhecimento médico. Entretanto, mesmo que especialistas em medicina, os profissionais de saúde não são especialistas em classificação. Ao buscar fontes de aquisição de conhecimento, é necessário analisar os critérios de qualidade conforme sugeridos por Uschold (1996, p.12 e 13), Gruber (1993, p.2 e 3), Seppälä e Ruttenberg (2013), Seppälä, Schreiber e Ruttenberg (2014), Köhler et al. (2006) e Spear (2006).

Para obter definições de leucemias foram usados vários dicionários especializados devido ao fato de haver definições pouco informativas ou sem características essenciais. No caso das leucemias, foi preciso recorrer a mais fontes de definições, já citadas, como os livros de patologia de Foucar et al. (2012), de hematologia de Hoffman et al. (2008), as classificações das leucemias da FAB e WHO, artigos científicos, dentre outras fontes. Para a pesquisa foram consultadas várias fontes de informação em saúde com a finalidade de analisar as definições existentes, abordagem esta que foi utilizada no método de formulação de definições.

As definições textuais das fontes pesquisadas revelam vários dos problemas descritos por Seppälä e Ruttenberg (2013), como por exemplo: definições circulares, intangíveis, uso de termos técnicos e múltiplas definições para o mesmo termo. Na presente pesquisa, por exemplo, o primeiro termo da amostra – *acute myeloid leukemia with myelodysplasia-related changes* – revelou circularidade; no segundo termo da amostra – *Acute myeloid leukemia with recurrent genetic abnormalities* – o problema principal a intangibilidade, devido à presença de termos técnicos. Para amenizar a intangibilidade, foi necessário explicar o significado das mutações genéticas para melhor compreensão da definição da leucemia.

A validação com o especialista de domínio do câncer foi essencial para definir a característica essencial de cada classe da LMA, uma vez que o oncologista possui o conhecimento necessário para definir cada tipo de leucemia. O especialista, ao analisar qual era a condição necessária e suficiente de cada classe da leucemia, usou a literatura médica para confirmar a característica essencial, principalmente quando haviam mais características. Pela validação, verificou-se que foi justamente a interferência do especialista que determinou a essência da definição, a qual foi confirmada com base nos critérios de diagnóstico da FAB e WHO. A última revisão do diagnóstico das leucemias foi realizada pela WHO, em 2008, ao reunir especialistas para rever as definições das LMAs. Entretanto, alguns tipos não foram revistos devido à sua raridade e à ausência de consenso sobre o diagnóstico.

A essência foi definida principalmente pelas características morfológicas, exceto pelo termo *acute myeloid leukemia with recurrent genetic abnormalities*, no qual a essência foi baseada nas características de anormalidades citogenéticas. Esse foi um termo cuja definição se mostrou mais complexa, uma vez que possui características que podem estar presentes em outros tipos de leucemias, além de apresentar definições circulares e intangíveis.

5 CONSIDERAÇÕES FINAIS

O método para a sistematização do processo, o qual se pretende denominar *OntoDef*, possibilita a criação de definições adequadas ao uso em ontologias, artefatos que fundamentam os modernos sistemas de informação. O experimento foi conduzido no domínio das leucemias e padronizado para contribuir com a teoria sobre definições em ontologias no âmbito da Ciência da Informação, bem como para melhorias das definições no domínio das leucemias. As definições das LMAs da pesquisa foram criadas com intuito de reutilização em outros sistemas e escritas

em língua estrangeira por se tratar de pesquisa internacional.

Definir o domínio das leucemias se mostrou uma tarefa árdua, complexa, que carece de especialistas e consome tempo. A complexidade terminológica das leucemias tornou o trabalho ainda mais laborioso, interferiu na aplicação do método proposto em relação à falta de consenso na definição das leucemias, além de exigir várias perspectivas científicas para o diagnóstico. A principal dificuldade ao definir foi encontrar a característica essencial das leucemias, uma vez que o domínio em si é um dos campos que estudam o câncer que mais apresenta diversidade de alterações fenóticas e genóticas no momento do diagnóstico.

O câncer é uma doença complexa, principalmente no caso das leucemias. Na atualidade, apenas uma característica morfológica – presença de 20% de blastos – não é suficiente para o diagnóstico e tratamento da doença. São várias características analisadas para definir a leucemia e, conseqüentemente, a conduta médica.

De fato, as LMAs são consideradas um dos cânceres mais complexos em relação ao diagnóstico e tratamento, apresentando altos índices de mortalidade. As classificações da FAB e WHO ainda apresentaram dificuldade ao categorizar, definir e diagnosticar os subtipos de LMA. Os esforços para melhor categorizar as neoplasias mieloides são muitos, como apontam Vardinan, Harry e Brunning (2002), que realizaram uma revisão sobre as classificações anteriores das neoplasias mieloides e leucemias agudas. Vardiman et al. (2009) publicou um estudo sobre a atualização da classificação da WHO, que além de usar a morfologia para definir as leucemias, acrescentou as informações de imunofenotipagem, citoquímica, genética e características clínicas. Tudo isso adicionou complexidade a busca pela essência de cada classe da leucemia considerando ainda que a ciência estuda atualmente formas de melhor categorizar este domínio.

Conhecer o domínio a ser definido é fundamental para aplicar o método de definições. No caso da pesquisa sobre leucemias, foi necessário estudar a patologia, o diagnóstico, a terminologia, a etiologia e discutir todas essas informações com o especialista oncologista, reafirmando a importância de conhecer o domínio.

O método é aplicável a outros domínios da Medicina e uma das principais contribuições desta pesquisa, além do exemplo prático de como elaborar definições bem formadas, é o esforço de sistematização de definição da característica essencial de cada tipo de leucemia. Como trabalhos futuros, espera-se automatizar a criação de definições formais, a partir das diretrizes da BFO.

Cabe a esse profissional da informação, que atua na área de saúde, desenvolver pesquisas sobre temas que envolvem ontologias, como instrumentos de representação para sistemas de informação, que podem contribuir para a padronização, interoperabilidade entre dados médicos, e comunicação entre as instituições de saúde. Cabe ressaltar que o trabalho da pesquisadora como bibliotecária junto à equipe médica da área de câncer foi essencial para o desenvolvimento do método *OntoDef*. A experiência contribuiu para aquisição de conhecimento sobre as leucemias, assim como para a validação das definições textuais.

Neste contexto, o bibliotecário e o profissional da informação têm o desafio de contribuir para o desenvolvimento de ontologias biomédicas, a partir de sua expertise na construção de tesouros, classificações, terminologias, sua habilidade na análise de assunto e na avaliação de fontes de informação. Definições bem construídas, fundamentadas em princípios universais, são o primeiro passo no sentido de obter comunicação e continuidade do atendimento adequado aos cidadãos, dada a atual prevalência dos sistemas de informação.

REFERÊNCIAS

ALMEIDA, Maurício Barcellos *et al.* A ontologia do sangue: uma iniciativa para representação e organização do conhecimento sobre Hematologia e Hemoterapia. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v.7, n.1, mar., 2013.

ALMEIDA, Maurício Barcellos *et al.* The Blood Ontology: an ontology in the domain of hematology. Status. In: INTERNATIONAL CONFERENCE OF BIOMEDICAL ONTOLOGIES, 2011, US. **Proceedings of the...** Disponível em: <http://ceur-ws.org/Vol-833/>. Acesso em: abr. 2014.

ARBER, Daniel A.; COUSAR, John B. Hematopoietic Tumors: Principles of Pathologic Diagnosis. Chapter 71. In: GREER, John P. *et al.* **Wintrobe's clinical hematology**. 13th. City: publication, 2013.

COELHO, Katia Cardoso. **Aquisição de conhecimento especializado para construção de ontologias**: um estudo no domínio das ciências da vida. 2012.174f. Dissertação (Mestrado em Ciência da Informação). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012. Disponível em: http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8XLQ54/disserta_o_katia_cardoso_coelho.pdf?sequence=2. Acesso em: ago. 2014.

DAHLBERG, Ingtraut. Teoria do Conceito. **Ci. Inf.**, Rio de Janeiro, v.7, n.2, p. 101-107, 1978a. (Trad. Astério Tavares Campos).

DAHLBERG, Ingtraut. A Referent-oriented analytical concept theory of interconcept. **International Classification**, v.5, n.3, p.142-150, 1978b.

DISEASE ONTOLOGY (DO). **Leukemia**. 2014. Disponível em: <http://disease-ontology.org/>. Acesso em: 29 set 2014.

FOUCAR, Kathryn *et al.* Myeloid neoplasm. In: FOUCAR, Kathryn.; *et al.* **Diagnostic pathology blood and bone marrow**. Monitoba:Amirsys, 2012. cap.9. 2-208.

GRUBER, Thomas. R. **Toward principles for the design of ontologies used for knowledge sharing**. Revision: August 23, 1993. Stanford Knowledge Systems Laboratory. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3207&rep=rep1&type=pdf>. Acesso em: 30 out 2013.

HOFFMAN, Ronald; *et al.* **Hematology**: basic principles and practice, 5th Ed. New York: Churchill Livingstone, 2008. p. 2560.

IWOOD. International Workshop on Definitions in Ontologies. 2014. In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGIES (ICBO), 5, 2014, **Proceedings...** Houston, TX, USA, October 7, 2014. Disponível em : <https://sites.google.com/site/definitionsinontologies/>. Acesso em: nov 2014.

- KEIZER, Nicolette F.; ABU-HANNA, Ameen.; ZWETSLOOT-SCHONK, Johanna.H. Understanding terminological systems. I: Terminology and typology. **Methods of information in medicine**, v.39, n.1, p.16-21. Mar. 2000.
- KÖHLER, Jacob *et al.* Quality control for terms and definitions in ontologies and taxonomies. **BioMed Central bioinformatics**, v.7, p. 212. Apr. 2006.
- MEEHAN, Terrence, F. *et al.* Logical development of the cell ontology. **BioMed Central bioinformatics**, v.12, p.6, 2011.
- MICHAEL, Joshua; MEJINO JUNIOR, Jose L.; ROSSE, Cornelius. The role of definitions in biomedical concept representation. *In: AMIA SYMPOSIUM, 2001. Proceedings...* [s.l.]; [s.n], 2001. p.463-7
- 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 (United States). **NCI dictionary of cancer terms: acute myeloid leukemia.** Disponível em : <http://www.cancer.gov/publications/dictionaries/cancer-terms?CdrID=46347>. Acesso em: 15 jul 2015.
- OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES (OBO) Foundry. **Home page.** Disponível em: <http://www.obofoundry.org/>. Acesso em: 20 maio de 2018.
- ONTOBEE. **Leukemia.** 2014a. Disponível em: http://www.ontobee.org/browser/rdf.php?o=DOID&iri=http://purl.obolibrary.org/obo/DOID_1240. Acesso em: maio 2014.
- ONTOBEE. **Welcome to Ontobee!** 2014b. Disponível em: <http://www.ontobee.org/>. Acesso em: 5 maio 2014.
- PETROVA, Alina; et al. Formalizing biomedical concepts from textual definitions. **Journal of biomedical semantics**. v.6, p.22. apr. 2015.
- SCHRIML, Lynn M.; et al. Disease Ontology: a backbone for disease semantic integration. **Nucleic Acids Research**. v.40. 2012. Disponível em: <http://nar.oxfordjournals.org/content/40/D1/D940.full.pdf?keytype=ref&ijkey=oOOVAj4zMzRpayq>. Acesso em: 21 set 2014.
- SEPPÄLÄ, Selja; SCHREIBER, Yonatan; RUTTENBERG, Alan. Textual and logical definitions in ontologies. *In: INTERNATIONAL WORKSHOP ON DRUG INTERACTION KNOWLEDGE MANAGEMENT (DIKR 2014), THE SECOND INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIES (IWOOD), 1., 2014, AND THE STARTING AN OBI-BASED BIOBANK ONTOLOGY WORKSHOP (OBIB 2014), CEUR Workshop Proceedings*, v.1309, Houston, TX, USA, October 6-7, p. 35-41. Disponível em: <https://seljaseppala.files.wordpress.com/2014/11/seppc3a4lc3a42014a.pdf>. Acesso em 21 maio 2015.

SEPPÄLÄ, Selja ; RUTTENBERG, Alan. 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. Disponível em: http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013_surveyreport_seppalaruttenberg.pdf. Acesso em: nov 2014.

SILVA, Grazielle C. da *et al.* Diagnóstico laboratorial das leucemias mieloides agudas. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 42, n. 2, p. 77-84, abr. 2006 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1676-24442006000200004&lng=en&nrm=iso. Acesso em: 26 ago. 2015.

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

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

SOERGEL, Dagobert. **Knowledge Organization Systems: Overview**.2014. Disponível em : <http://www.dsoergel.com/SoergelKOSOverview.pdf>. Acesso em: 14 ago 2014.

SPEAR, Andrew D. **Ontology for the twenty first century: An Introduction with Recommendations** .Institute for Formal Ontology and Medical Information Science (IFOMIS), Saarbrücken, Germany .University at Buffalo, Buffalo, New York, U.S.A. 2006. Disponível em : <http://ifomis.buffalo.edu/bfo/documents/manual.pdf>. Acesso em: 4 abril 2014.

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

THE GENE ONTOLOGY. **An introduction to the gene ontology**. Disponível em: <http://www.geneontology.org/>. Acesso em: 20 nov 2013.

TSATSARONIS, George; et al. Learning Formal Definitions for Biomedical Concepts. *In: SRINIVAS, K; JUPP, S. (ed.). Proceedings of the 10th OWL: Experiences and Directions Workshop (OWLED 2013)*, May 2013. Disponível em: <https://dill.inf.tu-dresden.de/web/LATPub509/en>. Acesso em: 21 maio 2015.

USCHOLD, Mike. **Building ontologies: Towards a Unified Methodology** (1996). Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.9075>. Acesso em: 15 fev 2013.

VARDIMAN, James.W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. **Blood**. v.114, n.5, p.937-51. jul. 2009.

XIANG, Zoushuang; et al. Ontobee: A Linked Data Server and Browser for Ontology Terms. *In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGIES (ICBO)*, 2nd. July 28-30, 2011, Buffalo, NY, USA. p.279-281. **Proceedings...**2011. URL: <http://ceur-ws.org/Vol-833/paper48.pdf>.



Artigo submetido ao sistema de similaridade

Submetido em: 11/10/2019 – Aceito em: 10/02/2020 – Publicado em: 12/05/2020
