



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento
Comunicação Oral

***ONTOFORINFOSCIENCE: UMA METODOLOGIA DETALHADA
PARA CONSTRUÇÃO DE ONTOLOGIAS E SUA APLICAÇÃO NO
DOMÍNIO DA BIOMEDICINA¹***

***ONTOFORINFOSCIENCE: A METHODOLOGY FOR BUILDING OF
ONTOLOGIES AND YOUR APPLICATION IN THE BIOMEDICAL
DOMAIN***

Fabício Martins Mendonça, UFMG
fabriciomendonca@gmail.com

Maurício Barcellos Almeida, UFMG
mba@eci.ufmg.br

Resumo. As ontologias, como instrumentos amplamente utilizados na estruturação do conhecimento biomédico, têm sido desenvolvidas através de metodologias diversas. Em geral, essas metodologias contêm passos bem estabelecidos, mas nem sempre bem detalhados, o que torna o desenvolvimento de ontologias um trabalho restrito a engenheiros do conhecimento. O presente artigo descreve uma metodologia para construção de ontologias, denominada *OntoForInfoScience*, que tem como diferencial o detalhamento das atividades do ciclo de desenvolvimento ontológico. O objetivo dessa nova metodologia é auxiliar especialistas em Organização do Conhecimento a superar problemas relativos ao jargão técnico e a questões lógicas e filosóficas no desenvolvimento de ontologias. De forma a identificar esses problemas, a *OntoForInfoScience* foi criada por cientistas da informação durante o desenvolvimento da HEMONTO, uma ontologia de domínio sobre os componentes do sangue humano para hematologia e hemoterapia. Apresenta-se aqui uma breve descrição da *OntoForInfoScience*, bem como os resultados práticos do desenvolvimento da ontologia sobre sangue. Conclui-se que a metodologia utilizada é de grande utilidade para a criação de representações ontológicas formais e representa uma iniciativa no sentido de possibilitar que mais cientistas da informação atuem de forma efetiva no desenvolvimento de ontologias, tendo em mente que para a realização desse fenômeno uma série de outras ações devem ser implementadas no contexto da Ciência da Informação. Como contribuição, além da metodologia em si, a pesquisa provê resultados parciais da ontologia no domínio do sangue.

Palavras-chave: Ontologias. Desenvolvimento de ontologias. Ontologias biomédicas. Hematologia.

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

Abstract. Ontologies, as instruments widely applied in the organization of biomedical knowledge, have been developed through several methodologies. In general, such methodologies contain well-established steps, but not always such steps are well detailed. Thus, it turns out that only knowledge engineers are able to perform all steps required to develop ontologies. This paper describes a methodology for ontology development, called *OntoForInfoScience*, which brings as novelty the detailing of each step of the ontology development cycle. The goal of this methodology is to make possible to experts in Knowledge Organization to overcome difficulties caused by technical jargon, as well as by logical and philosophical issues, which one can find when dealing with ontologies. In order to identify those difficulties and issues, *OntoForInfoScience* was created by information scientists along with HEMONTO, a domain ontology about blood components, which will be used in hematology and blood transfusion institutions. Here, we present a brief description of *OntoForInfoScience*, as well as the practical results in the development of the blood ontology. We conclude that the new methodology is useful for the creating formal ontological representations. As contribution, in addition to the methodology *per se*, our research also provides partial results of the development of HEMONTO.

Keywords: Ontologies. Ontologies development. Biomedical ontologies. Hematology.

1 INTRODUÇÃO

Ontologias têm sido amplamente utilizadas na representação e organização do conhecimento especializado das mais diversos campos da biomedicina. Especificamente, as ontologias são utilizadas na padronização dos vocabulários médicos como uma alternativa às terminologias médicas tradicionais (SMITH, 2004), uma vez que possuem maior expressividade para representação do conhecimento devido ao uso de formalismos lógicos. Além disso, ontologias possibilitam representar o conhecimento de forma mais próxima da realidade, o que é fundamental para a biomedicina (FREITAS; SCHULZ, 2009).

O aumento na aplicação de ontologias biomédicas pode ser notada, por exemplo, pelo surgimento de repositórios especializados e pela incorporação de ontologias na estrutura de fontes heterogêneas de conhecimento biomédico. O primeiro caso pode ser ilustrado pelo *Open Biomedical Ontologies (OBO)* (SMITH et al, 2007), um repositório de ontologias biomédicas disponíveis *online* em domínio público, que visa à integração e padronização. O segundo caso pode ser ilustrado pela formalização da *Unified Medical Language System (UMLS)* (BODENREIDER, 2004), que incluiu em sua estrutura a *Gene Ontology* (SMITH; WILLIAMS; SCHULZ, 2003) e a ontologia *Foundational Model of Anatomy (FMA)* (ROSSE; MEJINO, 2003).

O presente artigo descreve uma nova metodologia para construção de ontologias, denominada *OntoForInfoScience*, que tem como principal diferencial o detalhamento das atividades do ciclo de desenvolvimento ontológico (MENDONÇA, 2015). O objetivo dessa nova metodologia é auxiliar especialistas em Organização do Conhecimento, incluindo

cientistas da informação, a superar problemas relativos ao jargão técnico e a questões lógicas e filosóficas que envolvem o desenvolvimento de ontologias. Evidentemente, a solução de tais problemas no contexto da Ciência da Informação demanda muitas outras ações, como por exemplo, a discussão e inserção de requisitos de formação técnica-acadêmica nas disciplinas da área, as quais são necessárias, mas extrapolam o escopo da presente pesquisa. Ainda sim, a metodologia *OntoForInfoScience* representa uma iniciativa em direção a um maior entendimento de termos e detalhes técnicos (lógicos e filosóficos) do processo de desenvolvimento de ontologias por parte dos cientistas da informação.

A metodologia *OntoForInfoScience* em si foi criada por cientistas da informação durante o desenvolvimento de uma ontologia sobre os componentes do sangue humano para hematologia e hemoterapia. Tal ontologia, denominada de HEMONTO, está inserida no âmbito de um projeto de escopo maior – o *Blood Project*² – que tem por objetivo gerar uma linguagem formal sobre o sangue humano, considerando a carência atual de representações no domínio da hematologia e hemoterapia.

A partir do uso da metodologia *OntoForInfoScience* na construção da ontologia HEMONTO, obteve-se a formalização dos termos técnicos específicos do domínio do sangue. Nesse processo, a metodologia detalhou o emprego de ontologias de fundamentação, a criação de definições formais para classes (derivadas de definições textuais), a caracterização dos diferentes tipos de relações ontológicas, dentre outras questões técnicas. Estes detalhamentos se mostraram úteis para guiar o desenvolvedor de ontologias ao longo de toda a atividade.

O restante do presente artigo está organizado da seguinte forma. A segunda seção faz referência às ontologias biomédicas de domínio que, de alguma forma, representam e descrevem conceitos do domínio do sangue. A terceira seção apresenta referências as metodologias para construção de ontologias mais citadas e usadas atualmente, destacando brevemente as etapas da *OntoForInfoScience*. A quarta seção descreve a construção de parte da ontologia HEMONTO com o uso da nova metodologia e a quinta seção apresenta os resultados do desenvolvimento da ontologia. Por fim, a sexta seção traz considerações finais e conclusões sobre a pesquisa, bem como expectativas para trabalhos futuros.

2 ONTOLOGIAS BIOMÉDICAS E O DOMÍNIO DO SANGUE

A partir de pesquisa e análise nos atuais repositórios relacionados com a representação do domínio do sangue humano, verifica-se que o grau de cobertura desse domínio ainda é

² Disponível em: <http://mba.eci.ufmg.br/blo/>

insuficiente nas terminologias e ontologias biomédicas atuais. Além disso, verificou-se que quando os aspectos do sangue humano são representados em tais instrumentos, em sua maioria, não são representados com um grau de formalidade necessário. Apesar destes problemas, a referência às terminologias e ontologias biomédicas que descrevem aspectos do sangue é de grande importância para a construção da ontologia HEMONTO, uma vez que um dos princípios universalmente aceitos na área de engenharia ontológica é a reutilização de termos de ontologias existentes.

Essa seção apresenta uma revisão de literatura sobre ontologias biomédicas que tratam de aspectos relacionados ao sangue humano e, por esse motivo, representam uma fonte para reutilização de termos para a HEMONTO. É importante destacar também que a revisão de literatura não é exaustiva, mencionado apenas as principais e mais conhecidas ontologias biomédicas, e não todas as ontologias disponíveis na web. A Tabela 1 descreve essas ontologias através de seu nome e acrônimo, autores da ontologia, uma breve descrição e exemplos extraídos do conteúdo da ontologia que tem relação com o domínio do sangue.

Tabela 1: Ontologias com termos relacionados ao sangue

| Ontologia | Descrição | Exemplos relacionados |
|--|---|---|
| <i>Cell Ontology (CL)</i> (BARD, RHEE e ASHBURNER, 2005). | Ontologia que representa o conhecimento biomédico relativo aos tipos de células, incluindo células procariotas, células fúngicas, células animais e células vegetais. | hemal cell (blood cell), plasma cell; neutrophil, eosinophil and basophil; monocyte and T lymphocyte. |
| <i>Chemical Entities of Biological Interest (ChEBI)</i> (DEGTARENKO et al., 2008) | Dicionário ontológico de entidades químicas de interesse biológico. Está focada nos componentes químicos ou “moléculas pequenas”, assim chamadas em oposição as “macromoléculas”, tais como ácidos nucléicos e proteínas. | anticoagulant, water, nutrient, hormone |
| <i>Foundational Model Anatomy (FMA)</i> (ROSSE e MEJINO, 2003). | Ontologia sobre a anatomia humana. Contém classes anatômicas, como estruturas macromoleculares complexas e componentes de células. | taxonomia para a chamada <i>entidade anatômica do sangue</i> |
| GALEN (RECTOR et al., 2003) | Uma ontologia clínica que contém as definições da área de anatomia, além de conteúdos relacionados à fisiologia humana, patologia e sintomatologia – os sintomas. | Formalismos lógicos de anatomia, fisiologia e patologia relacionados ao sangue humano. |
| <i>Gene Ontology (GO)</i> (SMITH, WILLIAMS e SCHULZE-KREMER, 2003). | Ontologia que descreve produtos genéticos e suas funções em qualquer organismo. | celular component, fibrinogen complex, blood coagulation. |
| <i>Protein Ontology (PRO)</i> (NATALE et al., 2011). | Ontologia que representa o conhecimento sobre proteínas, complexos proteicos e as classes evolucionárias de proteínas, além das múltiplas formas de proteínas dos genes. | albumin, prothrombin, antithrombin-III; coagulation factors V, VII, VIII, IX, X, XI |

3 AS ETAPAS DA ONTOFORINFOSCIENCE

Além das ontologias biomédicas de domínio, outro tópico essencial para background

teórico da presente pesquisa diz respeito à metodologias para a construção de ontologias. São referenciadas metodologias de construção mais citadas na literatura da área, as quais são também as mais usadas no desenvolvimento de ontologias. Descreve-se ainda uma nova metodologia para construção de ontologias – *OntoForInfoScience* – que se propõe a detalhar as atividades envolvidas no ciclo de desenvolvimento ontológico, com a adoção de uma linguagem mais adequada e de fácil entendimento a desenvolvedores de ontologia que não sejam especialistas em Lógica, Computação e Filosofia (MENDONÇA, 2015).

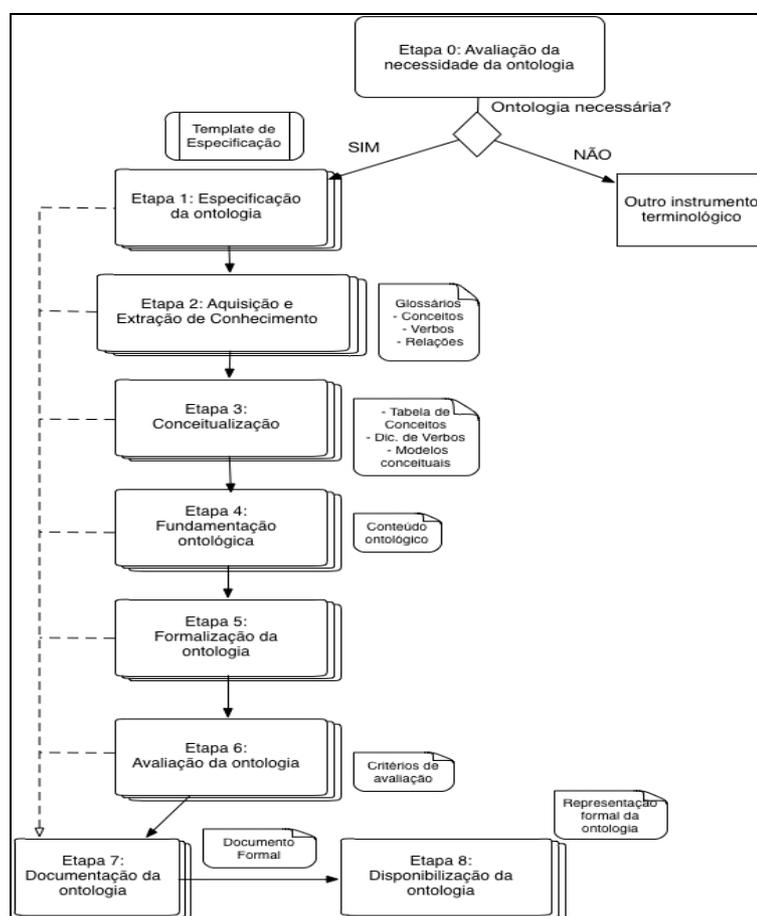
O uso frequente de ontologias para representação do conhecimento a partir dos anos de 1990, resultou no desenvolvimento de uma série de metodologias e métodos sobre como conduzir a construção de uma ontologia. Na literatura da área, é possível perceber uma proliferação de metodologias e métodos de construção entre meados da década de 90 até o início dos anos 2000. Nesse período surgiram as metodologias até hoje mais referenciadas por pesquisas na área, tais como: metodologia de Uschold e King (1995) – *Enterprise Ontology*; metodologia de Gruninger e Fox (1995) - *Toronto Virtual Enterprise (TOVE)*; a *Methontology* (GÓMEZ-PEREZ; FERNANDEZ-LOPES; VICENTE, 1996); o método 101 de Noy e McGuinness (2001); o método CYC de Reed e Lenat (2002); a *On-To-Knowledge Methodology (OTKM)* (SURE; STAAB; STUBER, 2003); dentre outras. De meados dos anos 2000 até hoje, também surgiram algumas metodologias para construção de ontologias que são bem referenciadas na área: a metodologia NeOn (SUARÉZ-FIGUEROA, 2008) e a *Up for ONtology (UPON)* (DE NICOLA; MISSIKOFF; NAVIGLI, 2009).

Mesmo com a disponibilidade de um número significativo de metodologias, diversos problemas relacionados ao seu uso permanecem em aberto. Um desses problemas é que tais metodologias não detalham suficientemente os passos que um desenvolvedor precisa executar ao longo do processo de construção. Em geral, as metodologias parecem considerar que o desenvolvedor já domina o assunto e não necessita de mais detalhes acerca de atividades e procedimentos envolvidos (MENDONÇA, 2015). Por exemplo, afirma-se que é necessário uma definição formal em alguma linguagem lógica para cada classe, porém, não se explica como proceder para criar definições formais, as quais exigem conhecimentos básicos em lógica. Além disso, o uso recorrente de conceitos técnicos da Ciência da Computação não elucidados adequadamente dificulta seriamente o trabalho de construção de ontologias por parte de outros profissionais (MENDONÇA, 2015). Exemplos desses conceitos são herança múltipla, cardinalidade, tipos de dados (*string*, inteiro, real, booleano), classe abstrata, classe disjunta, dentre diversos outros.

Considerando tais problemas, a *OntoForInfoScience* detalha os passos do ciclo de vida

ontológico em uma linguagem simples e adequada a profissionais diversos. Para tanto, essa nova metodologia reutiliza etapas de metodologias já existentes (por exemplo, *Methontology*, *101 Method* e *NeOn*), buscando suprir as limitações presentes em cada uma delas: detalham-se as etapas e passos reutilizados, além de realizar adaptações necessárias para obter uma linguagem apropriada a todos desenvolvedores de ontologia. A metodologia *OntoForInfoScience* engloba, ao todo, uma pré-etapa e oito etapas, conforme apresentado na Figura 1. No restante da presente seção apresenta-se uma descrição breve das etapas. A descrição completa está disponível em Mendonça (2015).

Figura 1: etapas da metodologia para desenvolvedores de Ciência da Informação



Na pré-etapa, o desenvolvedor realiza uma avaliação prévia sobre a real necessidade da construção de uma ontologia para o problema analisado. Se o objetivo desse desenvolvimento consiste, por exemplo, unicamente na indexação e recuperação de informações em um domínio específico, a construção de um tesauro é suficiente. Entretanto, se o objetivo consiste na representação de aspectos e objetos do mundo real e também a

necessidade do uso de relações mais extensíveis do que aquelas contidas em vocabulários controlados, a construção de uma ontologia é fundamental.

Decidida a necessidade de desenvolver a ontologia inicia-se a sequência de etapas do processo de construção. Na etapa 1 realiza-se a especificação da ontologia através do template de especificação, o qual deve conter informações sobre: o domínio e escopo, propósito geral, classes de usuários que representam o público-alvo da ontologia, cenários de aplicação para uso da ontologia e o grau de formalidade. Estabelece-se também o tipo da ontologia: ontologia de alto ou médio nível, de domínio ou de tarefa; ontologia leve ou pesada. Por fim, delimita-se do escopo de cobertura da ontologia descrevendo-se o ponto de partida, o limite do domínio coberto e questões de competência.

A etapa 2 consiste na aquisição e extração de conhecimento, que engloba as atividades de seleção dos materiais de referência do domínio e a adoção de métodos de extração do conhecimento. Na *OntoForInfoScience*, essas atividades são conduzidas de forma a combinar diferentes métodos de extração, como: análise textual e manual de textos, extração terminológica automática, método semiautomático de identificação de conceitos do domínio. Além disso, busca-se colaboração, isto é, a aquisição de conhecimento envolvendo diferentes visões do domínio a partir de consulta a especialistas diversos. Os artefatos resultantes da etapa 2 são glossários de conceitos, de verbos e de relações, os quais são utilizados como parâmetros de entrada na etapa 3.

Na etapa 3, de conceitualização, são realizadas atividades como identificação e análise dos conceitos do domínio que serão incluídos como classes da ontologia. Além disso, realiza-se a organização e a estruturação do conhecimento de forma a se obter relações, propriedades e restrições da ontologia. Na *OntoForInfoScience* esse processo ocorre a partir da transformação dos glossários construídos na etapa anterior em artefatos conceituais da etapa 3, os quais correspondem à: i) tabela de conceitos e propriedades, que contém os conceitos identificados e validados, suas definições, seus termos sinônimos, seus valores possíveis e suas propriedades (restrições) a nível conceitual; ii) dicionário de verbos, que contém os verbos identificados e validados como candidatos à relação na ontologia e suas definições textuais; e iii) os modelos conceituais gráficos, que representam as relações conceituais entre os conceitos identificados através de grafos ou estruturas similares.

A etapa 4 corresponde à fundamentação da ontologia desenvolvida, isto é, o desenvolvedor deve pesquisar por ontologias de fundamentação que atuem como ponto de partida e escolher aquela mais apropriada ao objetivo proposto. Ao fazer essa escolha, o desenvolvedor deve considerar a abordagem filosófica subjacente a ontologia de

fundamentação, de forma a justificar decisões ontológicas na modelagem do domínio. A ontologia de fundamentação selecionada deve ser importada para um editor de ontologias para início implementação.

Na etapa 5 realiza-se a representação formal da ontologia através de uma linguagem lógica. Nessa etapa produz-se a descrição formal do domínio a partir da conceitualização anterior (etapa 3). Portanto, o conhecimento do domínio, tratado anteriormente apenas a nível conceitual, passa a ser tratado a nível ontológico-formal, o que implica em uma série de adaptações nas estruturas conceituais de forma a atender as restrições ontológicas e formais. Estão incluídas como atividades: i) construir a taxonomia geral da ontologia, baseada na estrutura da taxonomia da ontologia de fundamentação escolhida; ii) definir propriedades descritivas das classes, as quais envolvem apenas atributos textuais, tais como nomes, sinônimos, definições textuais e anotações; iii) criar definições formais para cada uma das classes da ontologia usando uma linguagem lógica, de forma que a definição formal seja derivada da definição textual elaborada no passo anterior; iv) definir propriedades de dados das classes, as quais envolvem atributos como tipos de dados, cardinalidade e uso dos quantificadores universais ou existenciais; v) criar instâncias das classes da ontologia, de forma a testar o funcionamento da ontologia; vi) especificar relações ontológicas, o que consiste na aplicação de um conjunto definido de regras e princípios para a transição das relações conceituais para o nível formal-ontológico. Dentre essas regras, tem-se como exemplos: a restrição à inclusão de relações a apenas aquelas identificadas como entidades da realidade, a caracterização dos tipos de uma relação ontológica, tais como tipos *is_a* e tipos *part_of*, dentre outras; vii) definir propriedades das relações ontológicas, as quais incluem o nome da relação, a definição semi-formal, a propriedades lógicas básicas, o domínio e a faixa, dentre outras.

A etapa 6 envolve a avaliação da ontologia que, no caso da OntoForInfoScience, corresponde a um conjunto de critérios que permitem realizar tanto a validação ontológica (adequação da ontologia ao domínio – mundo real) quanto a verificação ontológica (análise da ontologia quando a corretude de sua construção). Exemplos de critérios de validação são: natureza ontológica, universalidade de classes e relações e grau de representatividade do domínio. Exemplos de critérios de verificação são: a não recursividade nas definições, a especificação dos diferentes tipos da relação *part_of*, a definição de relações inversas e criação de cardinalidades.

Na etapa 7, de documentação, produz-se o documento formal da ontologia. A elaboração da documentação é realizada, praticamente, ao longo de todo o processo de

construção. O conteúdo do documento formal da ontologia engloba o documento de especificação (etapa 1), materiais ou documentos de referência do domínio (etapa 2), conjunto de modelos conceituais (etapa 3), ontologias reutilizadas (etapas 4 e 5), conteúdo ontológico-formal (etapa 5), documento de avaliação da ontologia (etapa 6).

Por fim, a etapa 8 cuida para que a ontologia desenvolvida seja disponibilizada em meio eletrônico de fácil acesso e visualização aos usuários. Como passos desta etapa, tem-se: i) a exportação do conteúdo formal da ontologia (desenvolvido na etapa 5), através de algum editor de ontologias para alguma linguagem lógica, em geral OWL, para disponibilização na web; ii) a apresentação da ontologia em formato gráfico para mais fácil entendimento pelos usuários, o que pode incluir tanto um documento estático, quanto uma interface de busca ao conteúdo da ontologia.

4 A ONTOFORINFOSCIENCE NO DESENVOLVIMENTO DA HEMONTO

Com o propósito de obter uma representação formal dos componentes do sangue humano, além de testar a *OntoForInfoScience*, a ontologia HEMONTO foi desenvolvida seguindo as etapas e passos incluídos na nova metodologia. Essa seção descreve a construção da HEMONTO com base na metodologia adotada.

De acordo com a *OntoForInfoScience*, a etapa preliminar (pré-etapa) consistiu em avaliar a real necessidade do desenvolvimento de uma ontologia para representação formal dos componentes do sangue. A necessidade de um vocabulário único livre de ambiguidades, através de uma representação formal desse importante domínio médico, da forma mais próxima possível dos aspectos e entidades reais do sangue humano, justifica a construção da ontologia.

Na etapa seguinte, a etapa 1, especifica-se a ontologia. Dentre os elementos da especificação definidos nessa etapa destaca-se o propósito geral da HEMONTO, o qual define a ontologia como um instrumento de apoio a profissionais das áreas de hematologia e hemoterapia no exercício de sua prática clínica, explicitando-lhes as características de cada um dos produtos derivados do sangue, bem como dos processos utilizados na obtenção de tais produtos. Esse suporte visa informar aos profissionais sobre os procedimentos adequados na extração, manipulação e armazenamento dos componentes do sangue utilizados em hemoterapia. Além deste propósito geral, destaca-se também a delimitação do escopo da HEMONTO, o qual tem como ponto de partida entidades do mundo real, categorizados por meio das classes da ontologia BFO; e o termo “sangue” (*portion of blood*), importado da ontologia FMA. Assim, parte-se do termo “sangue” e, a partir de uma abordagem híbrida que

combina a estratégia *top-down* com a estratégia *bottom-up*, definem-se todos os componentes, tipos e derivados do sangue humano até o limite do domínio coberto pela HEMONTO. O limite nesse caso, foi definido com os componentes caracterizados como proteínas e enzimas. Daí para frente, faz-se referência a outras ontologias que definem termos, como a GO, a ChEBI e a Cell Ontology. A questão dos limites de escopo, bem como de granularidade, são sempre questões que geram dúvida na construção de ontologias.

Na etapa de aquisição e extração do conhecimento (etapa 2), utilizou-se como documentos de referência no domínio do sangue para construção da HEMONTO: i) um guia de hemocomponentes do Ministério da Saúde brasileiro; ii) o padrão terminológico ISBT 28 (ICCBBA, 2010); iii) o manual técnico sobre sangue e terapia celular da *American Association of Blood Banks* (AABB, 2005); iv) livro-texto sobre hematologia clínica *Wintrobe's Clinical Hematology* 12a edição (GREER et al., 2009). Tais documentos são considerados referência no domínio de acordo com avaliação de especialistas da área consultados. Quanto aos métodos de extração de conhecimento foram utilizados: i) o software de análise linguística SketchEngine³ para o processo de extração terminológica de termos frequentes encontrados nos documentos; ii) o framework colaborativo *Conceptualization Modelling Environment* (ConceptME) (SOUSA et al., 2013) para identificação e extração semi-automática de conceitos e relações do domínio; e iii) análise manual dos textos dos materiais de referência, por parte do desenvolvedor da ontologia, visando também a identificação de conceitos e relações candidatas.

Na etapa de conceitualização (etapa 3), o glossário de conceitos produzido na etapa 2 foi transformado, inicialmente, em um dicionário de conceitos, para depois ser finalmente transformado em uma tabela de conceitos e propriedades, com ajuda de especialistas no domínio. Essa tabela engloba os conceitos identificados e validados, suas definições, seus sinônimos, seus valores possíveis e suas propriedades (restrições) a nível conceitual. Ainda nessa etapa, o glossário de verbos foi transformado em um dicionário de verbos, contendo os verbos identificados e validados, bem como suas definições textuais; e o glossário de relações transformado em um conjunto de relações conceituais com apoio de especialistas. Por fim, todos esse material produzido na etapa 3 foram usados como referência no desenvolvimento de modelos conceituais gráficos do domínio do sangue utilizando as ferramentas *ConceptMe*⁴, *DiagramEditor*⁵ e *OmniGraffle*⁶.

³ Disponível em: <<http://Sketch Engine.co.uk/>>. Acesso em: 01 abr. 2015.

⁴ Disponível em: <http://www.conceptme.pt/conceptme/index.php/Main_Page>. Acesso em: 10 de dez. 2014.

⁵ Disponível em: <<http://dia-installer.de/>>. Acesso em: 10 dez. 2014.

Na etapa 4, fundamentação ontológica, foram adotadas duas ontologias de fundamentação: i) *Basic Formal Ontology* (BFO) (GRENON; SMITH, 2004), que foi usada como ponto de partida para a construção da HEMONTO e a definição das classes mais gerais; ii) *Relation Ontology* (RO) (SMITH et al., 2005), utilizada para especificação de relações ontológicas entre as classes do domínio. Ambas as ontologias estão inseridas no âmbito do realismo científico (BUNGE, 1974), mais especificamente em uma abordagem que tem sido chamada de realismo ontológico (GRENON; SMITH, 2004). Na aplicação das ontologias de fundamentação utilizadas, além de seus princípios formais seguidos ao longo do desenvolvimento, ambas as ontologias foram importadas com o uso do editor de ontologias Protégé 4.3.

A formalização do conteúdo ontológico da HEMONTO, atividade da etapa 5, foi realizada a partir dos seguintes passos: i) construção da taxonomia geral da ontologia, tomando como ponto de partida as entidades mais gerais da BFO e categorizando as entidades específicas do sangue abaixo delas; ii) definição das propriedades descritivas das classes junto ao fornecimento dos atributos textuais correspondentes no Protégé 4.3 (*ID, label, imported_from, hasSynonym, definition, comments, etc.*); iii) criação de definições formais em OWL-DL para cada classe da ontologia (vide exemplo para a definição da classe “*portion of plasma*” na Figura 2); iv) definição das propriedades de dados das classes para representar as características de cada componente do sangue, tais como: volume apropriado, porcentagem de hematócritos, temperatura de armazenamento, validade, entre outras; v) criação de instâncias das classes usando um conjunto de regras básicas, que envolvem a identificação de instâncias via questões de competência; vi) especificação de relações ontológicas a partir das regras e princípios para transição das relações conceituais para o nível formal-ontológico, tais como tipos de relações *is_a* e tipos de relações *part_of*; vii) definição das propriedades descritivas e lógicas de cada relação da ontologia, utilizando os atributos correspondentes no Protégé 4.3 (*label, definition, characteristics, inverse_of, domains, ranges, etc.*).

⁶ Disponível em: <<https://www.omnigroup.com/omnigraffle>>. Acesso em: 05 jan. 2015.

Figura 2: Exemplo de definição formal de classe na ontologia

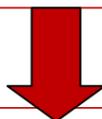
Portion of plasma: definição textual

Da FMA - Portion of plasma: portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.



Anotação da definição textual

Da FMA - Portion of plasma: portion of body substance [FMA/HEMONTTO: *portion of body substance*] corresponding to [BFO: *has_quality*] the liquid [FMA/HEMONTTO: *liquid*] component of [RO: *part_of*] blood [FMA/HEMONTTO: *portion of blood*] in which blood cells [FMA/HEMONTTO: *blood cell*] are immersed [RO: *contains*].



Portion of plasma: definição formal (em lógica OWL-DL)

Portion of plasma is a *portion of body substance* and (has_quality *liquid*) and (part_of *portion of blood*) and (contains *blood cell*).

Na etapa 6 foi realizada a avaliação do conteúdo ontológico da HEMONTTO com base no conjunto de critérios para validação e verificação ontológicas. Tais critérios abrangem aspectos relativos ao compromisso ontológico, especificação da ontologia, validação por especialistas, expansibilidade, completude, integridade, consistência, precisão e documentação. A documentação da HEMONTTO foi construída ao longo de todo processo de desenvolvimento e concluída na etapa 7, de forma a produzir um documento formal.

Por fim, na etapa 8, realizou-se a disponibilização da ontologia HEMONTTO. Inicialmente, para representação formal de todo o conteúdo desta ontologia optou-se pelo uso da linguagem OWL/XML, obtida através de exportação com o uso do software Protégé 4.3. Com relação a sua disponibilização para usuários da ontologia em geral, o recurso utilizado ontologia foi o uso do plug-in OWLDoc do Protégé, a fim de permitir uma visão geral e de fácil leitura para seus usuários de todas as estruturas ontológicas da HEMONTTO.

5 CONTEÚDO DA HEMONTTO

A partir da aplicação prática da *OntoForInfoScience* na construção da HEMONTTO foi possível gerar uma representação formal expressiva dos componentes do sangue humano para uso no cuidado à saúde da população. Essa seção descreve brevemente o conteúdo parcial da HEMONTTO desenvolvido ao longo do teste e criação da *OntoForInfoScience*.

A HEMONTTO é uma ontologia biomédica de domínio que, em sua versão atual, engloba um total de 209 termos, dos quais 155 são classes e 54 são relações. Seguindo a nomenclatura da BFO (GRENON; SMITH, 2004) 113 classes são continuantes e 41 classes são ocorrentes. A origem de cada uma dessas classes e relações da HEMONTTO é apresentada

na Tabela 2, a seguir.

Tabela 2: Ontologias de origem das classes da HEMONTO

| Classes | | Relações | |
|--|------|------------------------------------|------|
| Ontologia de origem | Núm. | Ontologia de Origem | Núm. |
| HEMONTA (específicas) | 62 | HEMONTA (específicas) | 14 |
| <i>Basic Formal Ontology</i> (BFO) | 34 | <i>Basic Formal Ontology</i> (BFO) | 20 |
| <i>Foundational Model Anatomy</i> (FMA) | 22 | <i>Relation Ontology</i> (RO) | 18 |
| <i>Protein Ontology</i> (PRO) | 14 | <i>Gene Ontology</i> (GO) | 2 |
| <i>Cell Ontology</i> (CL) | 9 | | |
| <i>Chemical Entities of Biological Interest</i> (ChEBI) | 4 | | |
| <i>Measurement Method Ontology</i> (MMO) | 4 | | |
| <i>Gene Ontology</i> (GO) | 4 | | |
| <i>Ontology for Biomedical Investigations</i> (OBI) | 1 | | |
| <i>Phenotypic Quality Ontology</i> (PATO) | 1 | | |
| <i>National Drug File Reference Terminology</i> (NDF-RT) | 1 | | |
| Total: | 155 | Total: | 54 |
| Total de entidades: 209 | | | |

As classes e relações da ontologia HEMONTA, citadas na Tabela 2, são descritas em detalhes no dicionário geral de classes HEMONTA e no dicionário de relações dessa ontologia. Ambos instrumentos incluem formalismos e axiomas lógicos para as entidades definidas na ontologia. A Tabela 3 abaixo mostra exemplos dessas entidades extraídas dos dicionários da HEMONTA.

Tabela 3: formalismos e axiomas lógicos das entidades da HEMONTA

| Entidade (classe ou relação) | Propriedades |
|---|--|
| <i>whole portion of blood</i> Sinônimos: <i>entire portion of blood; maximum portion of blood; portion of blood in whole body</i> | Importado de: FMA: <i>whole portion of blood</i> (FMAID: 263001) Definição textual: <i>Portion of blood collected into an anticoagulant and not further processed.</i> Definição formal: <i>is_a 'portion of blood' collected_in some anticoagulant and not (separated)</i> Comentários: <i>A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.</i> |
| <i>anticoagulant</i> Sinônimos: <i>anticoagulante, anticoagulants</i> | Importado de: CHEBI: <i>anticoagulant</i> (CHEBI: 50249) Definição textual: <i>An agent that prevents blood clotting.</i> Definição formal: <i>is_a realizable_entity and prevent_of some 'blood coagulation'</i> |
| <i>collected_into</i> | Origem: HEMONTA Definição semi-formal: C collected_into P – Relação ontológica identificada entre um continuante C e um ocorrente P, que é uma entidade realizável, tal que o continuante C é extraído ou coletado através de algum meio ou processo, que corresponde à entidade realizável P. Domínio: <i>material_entity</i> Imagem: <i>realizable_entity</i> Relação inversa: <i>has_collected</i> Exemplo de uso: |

| | |
|------------------------|---|
| | <i>whole portion of blood is_a 'portion of blood' collected_in some anticoagulant and not (separated)</i> |
| <i>participates_in</i> | <p>Origem: RO (RO_ID: 0056)</p> <p>Definição textual: <i>A relation between a continuant and a process, in which the continuant is somehow involved in the process.</i></p> <p>Definição formal: $\forall x,y(participates_in_{it}(x,y) \equiv mpart_of(x,y) \wedge ED(x) \wedge PD(y))$ onde ED é um endurente ou continuante e PD é um perdurante ou ocorrente.</p> <p>Relação inversa: <i>has_participant</i></p> <p>Domínio: <i>continuant</i></p> <p>Imagem: <i>occurrent</i></p> <p>Exemplo de uso: <i>plasma-extractors participates_in process of separation of the plasma</i> <i>this blood clot participates_in this blood coagulation</i></p> |

Além desses dicionários, o conteúdo da HEMONTO engloba estruturas de representação gráfica que mostram os relacionamentos entre as entidades do domínio. Tais estruturas são representadas por taxonomias, partonomias, e estruturas gerais ou mistas que envolvem mais de um tipo de relação entre as entidades. Taxonomias, partonomias e estruturas gerais de representação foram desenvolvidas tanto na etapa de conceitualização do domínio do sangue, dando origem aos modelos conceituais; quanto na etapa de formalização da ontologia, neste último caso, usando-se apenas relações ontológicas. As estruturas definidas na etapa de conceitualização compreendem: i) partonomia de uma porção de sangue; ii) taxonomia dos tipos de leucócitos; iii) partonomia dos produtos derivados do sangue; iv) partonomia dos tipos de análise do processo básico de amostras de sangue; v) taxonomia dos tipos de células sanguíneas patológicas; entre outras. Já as estruturas construídas na etapa de formalização englobam: i) uma taxonomia de entidades continuantes; ii) uma taxonomia de entidades ocorrentes; iii) taxonomia dos tipos de anti-coagulantes; iv) taxonomia dos parâmetros básicos do sangue; v) estrutura de representação do processo de punção venosa; vi) estrutura de representação dos processos de obtenção dos componentes concentrado de hemácias e concentrado de plaquetas; vii) estrutura de representação dos processos de obtenção dos componentes plasmáticos do sangue; viii) representação dos processos de obtenção do componente concentrado de granulócitos; dentre outras.

6 CONSIDERAÇÕES FINAIS

Um dos principais problemas que profissionais e pesquisadores em Ciência da Informação enfrentam no desenvolvimento de ontologias, é a presença de termos técnicos provenientes da computação, lógica e filosofia. Esse artigo apresentou a *OntoForInfoScience*, uma ontologia que tem como diferencial o detalhamento das atividades de construção de

ontologia além do que é apresentado em metodologia tradicionais. Dessa forma, pretende-se contribuir para expandir as possibilidades de desenvolvimento de ontologias, estendendo-a a profissionais que não são engenheiros do conhecimento. Apesar desse propósito com a presente pesquisa, sabe-se que uma solução mais abrangente e completa para as dificuldades dos cientistas da informação com o uso de termos técnicos e questões lógicas e filosóficas inerentes ao desenvolvimento de ontologias, passa necessariamente pela discussão e inserção de requisitos de formação técnica-acadêmica nas disciplinas relacionadas ao tema em Ciência da Informação. Trabalho este que é bastante amplo e ultrapassa o escopo desta pesquisa.

Na prática, a metodologia *OntoForInfoScience* foi utilizada e testada na construção de uma ontologia no domínio da hematologia e hemoterapia – a HEMONTO – e, nesse processo, mostrou-se eficiente quanto ao propósito de gerar uma representação formal adequada. Tal metodologia utiliza uma linguagem simples, que permite a compreensão das atividades executadas ao longo do processo de construção, sem perder a capacidade de expressividade do uso de linguagens lógicas. Algumas funcionalidades da *OntoForInfoScience* que são úteis e contribuem para o desenvolvimento de ontologias são: i) explicação sobre o significado e o preenchimento das propriedades descritivas e lógicas das classes e relações; ii) incentivo o uso de ontologias de fundamentação como ponto de partida do desenvolvimento ontológico; iii) detalhamento do processo de criação de definições textuais e formais das classes de uma ontologia; iv) explicação sobre o significado de uma relação ontológica e de que forma é possível caracterizar os tipos de relação ontológica para usá-los adequadamente na representação do domínio; v) incentivo a colaboração como um aspecto essencial na etapa de conceitualização, com o uso de ferramentas colaborativas; vi) explicação de formas para pesquisa por termos relacionados ao domínio para importação na ontologia. Considerando tais funcionalidades e sua aplicabilidades prática, a metodologia *OntoForInfoScience* foi de grande utilidade para os propósitos de desenvolvimento da HEMONTO. Por limitações de espaço, o aspecto mais importante da *OntoForInfoScience*, a saber, o detalhamento das atividades, não pôde ser apresentado em sua totalidade aqui. Para uma visão completa consulte Mendonça (2015).

Considerando a carência de uma representação formal do sangue e seus componentes, justifica-se a pesquisa sobre a ontologia biomédica HEMONTO. Essa ontologia contribui para a descrição dos componentes do sangue e da forma como devem ser extraídos, manipulados e armazenados para assegurar eficiência no cuidado à saúde dos pacientes. Para que se obter a representação formal expressiva e abrangente do domínio, atualmente não encontrada nas terminologias e ontologias biomédicas que fazem referência ao sangue humano, a

HEMONTO foi desenvolvida a partir da nova metodologia para construção de ontologias. Nesse sentido, um dos objetivos da presente pesquisa foi mostrar como se dá o processo de construção da ontologia HEMONTO a partir dessa nova metodologia e os resultados alcançados.

Quanto ao conteúdo da HEMONTO e os resultados parciais, pode-se afirmar que a ontologia desenvolvida caracteriza-se por conter uma representação formal mais abrangente no domínio do sangue do que aquela disponível nas terminologias biomédicas, tais como o FMA, GALEN, UMLS, GO, dentre outras. Ainda assim, a HEMONTO é uma ontologia em desenvolvimento, devido ao seu amplo escopo de cobertura. Sabe-se que o processo de construção de uma ontologia é um processo iterativo e sua validação em diferentes comunidades de especialistas, diferentes corpos médicos, é fundamental para avaliação de sua utilidade prática. Nesse sentido, a HEMONTO precisa de uma maior validação pós-desenvolvimento entre especialistas da área, além de revisões e atualizações em seu conteúdo ontológico.

REFERÊNCIAS

AMERICAN ASSOCIATION OF BLOOD BANKS (AABB) 17th edition. **Technical Manual**. Bethesda, Maryland: AABB, 2011.

BARD, J.; RHEE, S. Y.; ASBURNER, M. An ontology for cell types. **Genome Biology**, v. 6, n. R21, 2005.

BODENREIDER, O. **The Unified Medical Language System (UMLS): integrating biomedical terminology**. Oxford University. 32(1) Suppl.1: D267- D270, 2004.

BUNGE, M. **Teoria e realidade**. São Paulo: Editora Perspectiva, 1974.

DE NICOLA, A.; MISSIKOFF, M.; NAVIGLI, R. A software engineering approach to ontology building. **Information Systems**, n. 34, p. 258–275, 2009.

DEGTYARENKO, K.; et al. ChEBI: a database and ontology for chemical entities of biological interest. **Nucleic Acids Research**, v. 36, D344–D350, 2008.

FREITAS, F.; SCHULZ, S. Pesquisa de terminologias e ontologias atuais em biologia e medicina. **RECIIS – R. Eletr. de Com. Inf. Inov. Saúde**. Rio de Janeiro, v. 3, n. 1, p. 8-20, mar. 2009.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. Towards a method to conceptualize domain ontologies. In: ECAI WORKSHOP ON ONTOLOGICAL ENGINEERING, 1996, Budapest. Disponível em: <http://citeseer.ist.psu.edu/483876.html>. Acesso em: 13 mar. 2014.

GRENON, P.; SMITH, B. SNAP and SPAN: towards dynamic spatial. **Spatial Cognition & Computation**, v. 4, n. 1, p. 69-104, 2004. Disponível em: http://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf. Acesso em: 12 abr. 2012.

GREER, J. P.; et al. **Wintrobe's Clinical Hematology 12th Edition**. Philadelphia: Lippincott Williams & Wilkins, 2009.

GRUNINGER, M.; FOX, M. S. “**Methodology for the design and evaluation of ontologies**”. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING (IJCAI). [S.l.; s.n.], 1995.

ICCBBA – ISBT 128 Standard. **Standard Terminology for Blood, Cellular Therapy and Tissue Product Descriptions**, v. 3.33, January 2010. Disponível em: <www.iccbba.org>. Acesso em: 8 jan. 2014.

IHTSDO-SNOMED – International Health Terminology Standards Development Organisation Systematized Nomenclature of Medicine, 2012. Disponível em: <<http://www.ihtsdo.org/snomed-ct/>>. Acesso em: 05 maio 2012.

MENDONÇA, F. M. **OntoForInfoScience**: metodologia para construção de ontologias pelos cientistas da informação – uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). Tese (Doutorado)-Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2015.

NATALE, D. A. et al. The protein ontology: a structured representation of protein forms and complexes. **Nucleic Acids Research** 2011, v. 39, 2011.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

OMS – Organização Mundial de Saúde. Disponível em: <<http://new.paho.org/bra/>>. Acesso em: 13 set. 2012.

RECTOR, A.; ROGERS, J. E.; ZANSTRA, P. E.; HARING, E. OpenGALEN: open source medical terminology and tools. In: AMIA ANNUAL SYMPOSIUM, 2003. **Proceedings...** 2003.

REED, S. L.; LENAT, D. B. **Mapping ontologies into cyc**, 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 25 abr. 2002.

ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: the Foundational model of anatomy. **Journal of Biomedical Informatics**, v. 36, p. 478–500, 2003.

SMITH, B. **Ontology and information systems**, 2004. Disponível em: <<http://www.ontology.buffalo.edu/ontology>>. Acesso em: 16 mar. 2012.

_____ et al. Relations in biomedical ontologies. **Genome Biology**, n. 6, R46, 2005.

_____ et al. The OBO foundry: coordinated evolution of ontologies to support biomedical dataintegration. **Nature Biotechnology**, v. 25, n. 11, p. 1251–1255, 2007.

_____ ; CEUSTERS, W. Ontological realism: a methodology for coordinated evolution of scientific ontologies. **Applied Ontology**, n. 5, p. 139–188, 2010.

SOUSA, C.; PEREIRA, C.; SOARES, A. Collaborative elicitation of conceptual representations: a corpus-based approach. **Advances in Information Systems and Technologies**, v. 206, p. 111-124, 2013,

SUÁREZ-FIGUEROA, M. C.; AGUADO et al. **NeOn D5.4.1. NeOn Methodology for Building Contextualized Ontology Networks**. NeOn project. Disponível em: <<http://www.neon-project.org>>. Acesso em: fev. 2008.

SURE, Y.; STAAB, S.; STUBER, R. **On-To-Knowledge Methodology (OTKM)**. 2003.

Disponível em:

http://www.sfu.ca/~mhatala/iat881/papers/2003_ontohandbook_otkmethodology.pdf.

Acesso em: 17 mar. 2014.

USCHOLD, M.; KING, M. **Towards a methodology for building ontologies**. 1995.

Disponível em: <http://citeseer.ist.psu.edu/uschold95toward.html>. Acesso em: 11 mar. 2014.