



Um estudo de caso na construção de ontologias biomédicas: uma ontologia de domínio sobre hemoterapia

Fabrizio M. Mendonça¹, Maurício B. Almeida¹

¹Escola de Ciência da Informação – Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627 - Campus Pampulha – 31.270-901 – Belo Horizonte – Brasil

fabriciomendonca@gmail.com, mba@eci.ufmg.br

Resumo. *O uso de ontologias possibilita organizar, explicitar e compartilhar o conhecimento de um dado domínio. O presente artigo descreve um estudo de caso no domínio da hemoterapia, em que se tem construído uma ontologia sobre componentes do sangue humano. Como pesquisa em andamento, apresentam-se os resultados parciais obtidos com a ontologia proposta.*

Abstract. *The use of ontologies enables one to organize, to make explicit and to share knowledge about a certain domain. This article describes a case study in the domain of blood transfusion, in which we have developed an ontology about human blood components. As an ongoing research, we present the partial results obtained with the proposed ontology.*

1. Introdução

Nos últimos anos, ontologias têm sido amplamente utilizadas na descrição formal do conhecimento científico para uso em sistemas de informação. O uso de ontologias representa uma evolução nas práticas atuais de modelagem para sistemas de informação, uma vez que possibilitam explicitar o conhecimento de um domínio, promovem o compartilhamento do conhecimento e favorecem a integração da informação [Guarino 1998], [Wand e Weber 2004].

Nesse contexto, o presente artigo descreve um estudo de caso no domínio do sangue humano que envolve a construção de uma ontologia sobre hemocomponentes e hemoderivados, denominada de HEMONTO no âmbito de um projeto biomédico de escopo maior denominado de *Blood Project* [Almeida et al. 2011]. Planeja-se que a ontologia resultante funcione como um repositório de conhecimento científico, sendo utilizada, por exemplo, como instrumento para anotação e, nesse sentido, preencher a lacuna causada pela falta de uma representação formal voltada para os hemoderivados do sangue e pelas possibilidades limitadas de recuperação da informação dessa área em função do uso de ferramentas gerais. Além disso, espera-se que a ontologia proposta possa facilitar as atividades de modelagem ou avaliação dos sistemas de informação no domínio tratado, podendo assim ser classificada como uma *ontology-driven information systems* [Guarino 1998], baseada em um conjunto de princípios que se convencionou chamar de realismo ontológico [Smith and Ceusters 2010].

O restante do presente artigo está estruturado da seguinte forma: a seção 2 apresenta a metodologia adotada para a construção da ontologia; na seção 3, apresenta-se o conteúdo parcial da HEMONTO; e, na seção 4, são feitas as considerações finais.

2. Metodologia de construção da ontologia proposta

A construção de tal ontologia HEMONTO foi conduzida da seguinte forma: (i) aquisição de conhecimento no domínio em estudo através de materiais de referência; (ii) reaproveitamento do conteúdo de outros glossários, padrões e outras ontologias; e (iii) utilização de alguns artefatos de representação para explicitar o conhecimento representado pela ontologia. Os passos (i) e (ii) estão descritos na subseção 2.1 e o passo (iii) é descrito na subseção 2.2.

2.1 Ontologias e materiais de referência

Para o estudo do domínio tratado, foi selecionado inicialmente guia para o uso de hemocomponentes [Brasil 2008]. O documento aborda diretrizes para manipulação de hemocomponentes, funcionando como um instrumento de apoio para a prescrição médica referente à escolha do hemocomponente mais adequado para transfusão.

Paralelamente ao estudo do domínio tratado, procedeu-se com uma revisão de literatura de ontologias relevantes para o domínio do sangue e domínios correlatos, uma vez que se desejava reaproveitar termos e relações. Em geral, desenvolve-se a ontologia de domínio a partir da hierarquia de uma ontologia de alto nível, e reaproveita-se termos de outras ontologias de domínio. Para desenvolvimento da HEMONTO, as ontologias selecionadas até o momento foram: (i) *Basic Formal Ontology* (BFO) [Grenon and Smith 2004], para a definição de classes genéricas; (ii) *Relation Ontology* (RO) [Smith et al. 2005], para definição de relações; e (iii) *Foundational Model of Anatomy* (FMA) [Rosse and Mejino 2003], para definição de parte das classes específicas de domínio. No restante desse artigo, apresentam-se as entidades em itálico e as relações em negrito.

2.2 Artefatos de representação: entidades e relações

Na construção da HEMONTO, os seguintes artefatos de representação têm sido utilizados: (i) *taxonomias*, correspondendo à relação formal **is_a**; (ii) *partonomias*, correspondendo a relação formal **part_of**; (iii) *relações não-hierárquicas e outras*, tais como: **participates_in**, **has_agent**, **produces**, **has_quality**; e (iv) uso de uma *sintaxe semi-formal* para especificar as relações ontológicas, conforme recomendado pelo repositório *Open Biomedical Ontologies* (OBO).

Tendo como base o padrão notacional da sintaxe adotada pela OBO, foram utilizadas 9 relações ontológicas da RO para ligar os termos da HEMONTO. Dessas 9 relações, sete são usadas nas estruturas de representação apresentadas neste artigo, e podem ser assim definidas formalmente [Smith et al. 2005]:

- (1) $C \text{ is_a } C1: \forall c, \forall t$, se $c \text{ instance_of } C$ at t então $c \text{ instance_of } C1$ at t , tal que $C, C1$ suportam apenas entidades continuantes e $c \text{ instance_of } C$ at t é uma relação primitiva de instanciação, na qual a entidade continuante particular c instancia o universal C num dado tempo t .
- (2) $P \text{ is_a } P1: \forall p$, se $p \text{ instance_of } P$ então $p \text{ instance_of } P1$, tal que $P, P1$ suportam apenas entidades ocorrentes e $p \text{ instance_of } P$ é uma relação primitiva de instanciação, na qual a entidade ocorrente particular p instancia o universal P .
- (3) $C \text{ part_of } C1: \forall c, \forall t$, se $c \text{ instance_of } C$ at t então há algum $c1$ tal que $c1 \text{ instance_of } C1$ at t e $c \text{ part_of } c1$ at t , onde $c \text{ part_of } c1$ at t é uma relação primitiva entre dois particulares continuantes, na qual um é parte do outro no tempo t mencionado.
- (4) $C \text{ participates_in } P: \forall c, \forall t$, se $c \text{ instance_of } C$ at t então há algum p tal que $p \text{ instance_of } P$ e $p \text{ has_participant } c$ at t , onde **has_participant** c at t é uma relação primitiva entre um processo e um continuante num tempo t , tal que o continuante participa do processo de alguma maneira.

- (5) P produces C : $\forall p$, se p instance_of P então há algum c e algum t ; tal que se c instance_of $C1$ at t and p produces c at t , onde p produces c at t é uma relação entre o processo p , o continuante c e um tempo t , no qual p produces c se algum processo que occurs_in p has_output c .
- (6) P preceded_by $P1$: $\forall p$, se p instance_of P então há algum $p1$ tal que $p1$ instance_of $P1$ and p preceded_by $p1$, onde p preceded_by $p1 = \forall t, t1$, se p occurring_at t and $p1$ occurring_at $t1$, então $t1$ earlier t , onde t earlier $t1$ é uma relação primitiva entre dois tempos tal que t ocorre antes de $t1$ e p occurring_at $t = \text{for some } c, p$ has_participant c at t .
- (7) C has_quality Q : é uma relação entre um continuante C e uma qualidade Q , a qual C has_quality Q se somente se: $\forall c, \forall t$, se c instance_of C at t então há algum $c1$ tal que: se $c1$ instance_of $C1$ and exists $\forall q, \forall t$, se q instance_of Q at t então há algum $q1$ tal que: se $q1$ instance_of $Q1$, tal que q inheres_in c at t .
- (8) Q is_quality_measured_as q : é uma relação entre um universal continuante Q e um particular continuante q , sendo que ambos são qualidades e $\forall q, \forall t$, se q instance_of Q at t então há algum $q1$ tal que $q1$ instance_of $Q1$.

Sobre essas definições formais das relações apresentadas, cabe ressaltar que: (i) as variáveis $c, c1, C$ e $C1$ representam entidades continuantes; (ii) $p, p1, P$ e $P1$ entidades ocorrentes; (iii) $q, q1$ e Q representam qualidades; (iv) $t, e t1$ representam instâncias de tempo; e (v) variáveis minúsculas, tais como p e c , correspondem a particulares e variáveis maiúsculas, tais como P e C , correspondem a universais.

3. Resultados

A HEMONTO representa o conhecimento sobre hemocomponentes e hemoderivados, englobando elementos constituintes e os procedimentos empregados para sua obtenção. Em sua versão atual, a ontologia possui 53 termos, dos quais 44 são classes da ontologia e 9 são relações. Dentre as classes, trinta e três classes são específicas da ontologia, 9 classes foram importadas da BFO e 2 classes da FMA. As 9 relações são importadas da RO. O editor de ontologias Protege 4.2¹ foi utilizado para a construção da ontologia e possibilitou a sua implementação na linguagem *Ontology Web Language* (OWL). Para a construção dos diagramas da ontologia proposta foi utilizado o software Diagram Editor². No restante dessa seção, descreve-se o conhecimento utilizado para criar a ontologia e apresentam-se diagramas.

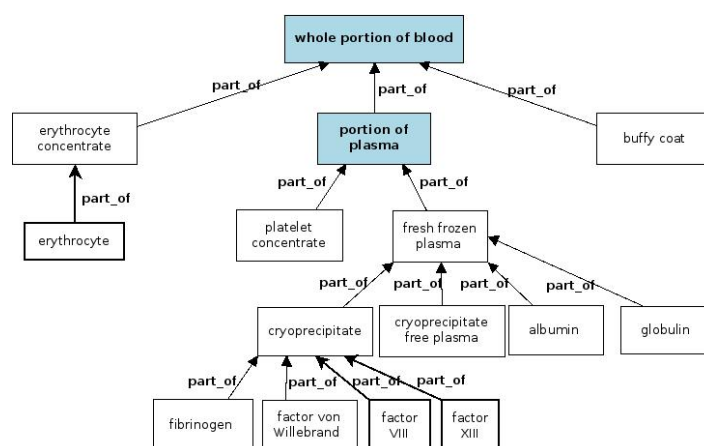


Figura 1: Partonomia dos componentes do sangue humano.

O sangue (*FMA:portion of blood*) é o líquido principal do corpo humano, formado de *plasma* (*FMA:portion of plasma*) e células sanguíneas. Para obter

¹ Disponível em: <http://protege.stanford.edu/>. Acesso em: 03 de Setembro de 2013.

² Disponível em: <https://projects.gnome.org/dia/>. Acesso em: 03 de Setembro de 2013.

hemocomponentes e hemoderivados do sangue é necessário submeter uma unidade de *sangue total* (FMA:whole portion of blood) a processos de centrifugação e congelamento. De acordo com o FMA, o sangue total é um tipo de uma porção de sangue, cujos seus componentes não foram separados. Esses tipos de hemocomponentes e hemoderivados estão representados em uma partonomia (veja figura 1).

Na figura 1, os retângulos sombreados representam as entidades FMA:whole portion of blood e FMA:portion of plasm. O whole portion of blood é submetido à centrifugação e separado inicialmente em três produtos: (i) erythrocyte, que armazenado em certas condições de temperatura gera o hemocomponente erythrocyte concentrate; (ii) FMA:portion of plasma, que corresponde ao plasma em sua situação natural (bruta), ainda rico em plaquetas (termo sinônimo: platelet rich plasma); e (iii) buffy coat, porção do sangue formada por leucócitos e plaquetas. Após um novo processo de centrifugação em alta rotação, a portion of plasma é separado em: (i) hemocomponente platelet concentrate e (ii) o fresh frozen plasma, com baixa porcentagem de plaquetas. O fresh frozen plasma pode ser submetido a um processo de extração de um de seus componentes – o cryoprecipitate – dando origem a dois outros hemocomponentes: (i) o próprio cryoprecipitate e (ii) o cryoprecipitate-free plasma. Do fresh frozen plasma ainda é possível extrair dois hemoderivados (i – albumin e ii – globulin), a partir do fracionamento desse plasma em processos industriais. Por fim, o cryoprecipitate contém uma série de glicoproteínas de alto peso molecular (fibrinogen, factor Von Willebrand, factor VIII e factor XIII) que cumprem o papel de fatores de coagulação na hemoterapia. A partir dessas proteínas é possível gerar o hemoderivado clotting factors concentrate.

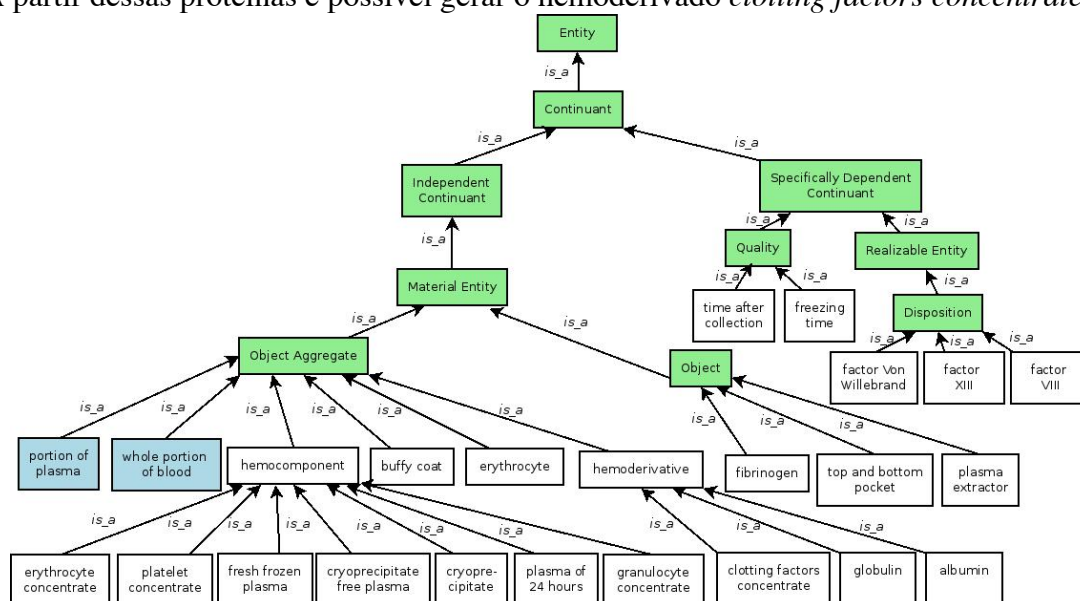


Figura 2: Taxonomia das entidades continuantes da HEMONTO.

De acordo com a estrutura taxonômica da ontologia de alto nível adotada, a BFO, tem-se dois grandes grupos de entidades: (i) *continuentes*, entidades que persistem ao longo do tempo mantendo sua identidade e que não possuem partes temporais; e (ii) *ocorrentes*, entidades que se revelam, se manifestam, ou se desenvolvem ao longo do tempo e possuem partes temporais [Spear 2006]. Seguindo tal estrutura, foram criadas duas taxonomias das entidades incluídas na HEMONTO. A

taxonomia da figura 2 representa o conjunto de *entidades continuantes* da HEMONTO, na qual as entidades importadas da BFO e do FMA estão sombreadas.

A entidade *FMA:whole portion of blood is_a BFO:object aggregate*, da mesma forma que seus componentes iniciais (*FMA:plasma, buffy coat and erythrocyte*). São também classificados como *BFO:object aggregate* os *hemocomponents* e os *hemoderivatives* e seus tipos específicos: (i) *erythrocyte concentrate, platelet concentrate; fresh frozen plasma; cryoprecipitate free plasma, cryoprecipitate, plasma of 24 hours* e *granulocyte concentrate*; e (ii) *clotting factors concentrate, globulin* e *albumin*. As entidades *plasma extractor, top and bottom pocket* e a proteína *fibrinogen* são classificadas como *BFO:object*, e os fatores de coagulação *factor Von Willebrand, factor VIII* and *factor XIII*, são classificados como *BFO:disposition*. Por fim, as entidades *time after collection* and *freezing time* foram classificadas como *BFO:quality*.

O outro grande grupo de entidades da HEMONTO corresponde aos *ocorrentes*, as quais representam os processos para obtenção dos hemocomponentes: *process of freezing; process of defrosting; process of centrifugation, process of centrifugation at high rotation; process of collection; extraction of buffy coat; remotion of plasm* and *remotion of cryoprecipitate*. Todas essas entidades foram classificadas como *BFO:process*. Por limitações de espaço, os processos não são representados na figura.

O plasma é um dos mais importantes componentes do sangue humano e, a partir dele, são gerados quatro hemocomponentes do sangue, conforme mostrado na figura 3, a seguir. Os retângulos da figura 3 representam classes da ontologia e as elipses representam propriedades destas classes.

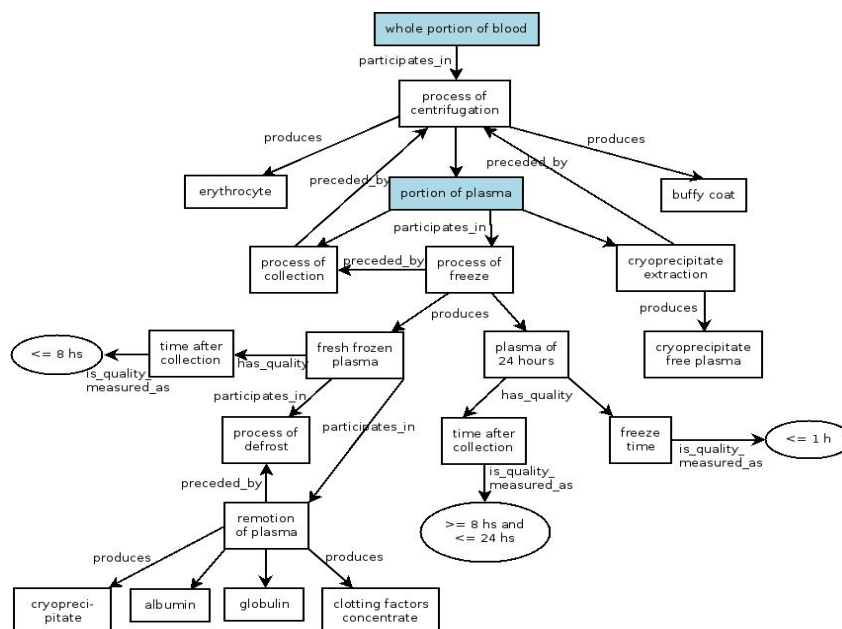


Figura 3: Processos de obtenção dos componentes plasmáticos

Para obtenção dos hemocomponentes *fresh frozen plasma* e *plasma of 24 hours* (lado esquerdo, figura 3), o procedimento inicial é o *process of centrifugation* do *whole portion of blood* para a separação de *erythrocyte, buffy coat* e *portion of plasma*. O passo seguinte consiste no *process of collection* da porção de plasma obtida. O tempo decorrido após a coleta do plasma, denominado de *time after collection*, é um parâmetro importante no processo, pois determina qual hemocomponente será obtido: (a) se o

tempo é no máximo 8 horas obtém-se o *fresh frozen plasm* e (b) quando o tempo pós-coleta está entre 8 horas e 24 horas obtém-se o *plasma of 24 hours*. Para que esses hemocomponentes sejam gerados é necessário que após o processo de coleta, o plasma seja submetido ao *process of freezing*. No caso do *plasma of 24 hours*, o tempo de congelamento deve ser no máximo de 1 hora, ou seja, *plasma of 24 hours has_quality freeze time is_quality_measured_as* <= 1 h. Ainda nesse diagrama, representa-se a obtenção do *cryoprecipitate free plasma* e do próprio *cryoprecipitate*. Para obter o *cryoprecipitate free plasma* (lado direito, figura 3), a etapa inicial corresponde novamente ao *process of centrifugation* do *whole portion of blood*, em seguida, tem-se o processo de *cryoprecipitate extraction* do plasma. Após essa extração obtém-se o *cryoprecipitate free plasma*. Já o *cryoprecipitate* é obtido a partir do *fresh frozen plasm* (lado esquerdo, figura 3) à temperatura de 1° C a 6° C. Esse plasma é submetido ao *process of defrosting* e, em seguida, o plasma sobrenadante é removido (*remotion of plasma*), deixando-se na bolsa coletora apenas a proteína precipitada e 10-15 ml deste plasma. Esses produtos formam o *cryoprecipitate*. No processo de *remotion of plasma*, o plasma removido **produces** *albumin, globulin e clotting factors concentrate*.

4. Considerações finais

O presente artigo apresentou um estudo de caso no domínio do sangue humano através da descrição do conteúdo parcial de uma ontologia sobre hemocomponentes e hemoderivados do sangue. A etapa seguinte desta pesquisa consiste na validação do conteúdo da ontologia, por parte de profissionais especialistas e a incorporação de novos termos à ontologia. A fim de possibilitar a validação do conteúdo da HEMONTO por parte dos especialistas, será criada uma interface de busca aos termos da ontologia, disponibilizada na web.

Referências

- Almeida, M. B.; Proietti, A. B.; Smith, B. and Ai, J. (2011). "The Blood Ontology: an ontology in the domain of hematology". In: *ICBO 2011*; Buffalo, USA.
- Brasil. Ministério da Saúde(2008). "Guia para o uso de hemocomponentes". Brasília, DF.
- Grenon, P.; Smith, B. (2004). "SNAP and SPAN: Towards Dynamic Spatial". *Spatial Cognition & Computation*, v.4, n.1, p. 69-104.
- Guarino, N. (1998). "Formal Ontology and Information Systems". In: *FOIS'98*; november 20, 2007; Trento, Italy. Edited by Guarino, N. IOS Press; pp. 3-15.
- Rosse, C.; Mejino, J. L. V.(2003). "A reference ontology for biomedical informatics: the Foundational Model of Anatomy". *Journal of Biomedical Informatics*, 36:478-500.
- Smith, B.; Ceusters, W.; Klagges, B.; Köhler, J.; Kumar, A.; Mungall, C.; Rector, A. L.; Rosse, C. (2005). "Relations in biomedical ontologies". *Genome Biology* 6, R46.
- Smith, B.; Ceusters, W. (2010). "Ontological realism: A methodology for coordinated evolution of scientific ontologies". *Applied Ontology* 5, p. 139–188. IO Press.
- Spear, A. D. (2006). "Ontology for the twenty first century: an introduction with recommendations". Saarbrücken, Germany.
- Wand, Y.; Weber, R. (2004). "Reflection: Ontology in Information Systems". *Journal of Database Management* 15 (2), iii-vi.