

Um estudo de caso para aquisição de conhecimento no domínio da hematologia

Kátia C. Coelho¹, Mauricio B. Almeida², Viviane Nogueira³

¹ Fundação Centro de Hematologia e Hemoterapia de Minas Gerais (Hemominas)
R. Grão Pará, 882 – Santa Efigênia – 30130110 – Belo Horizonte – Brasil

^{2,3} Escola de Ciência da Informação – Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627 – Campus Pampulha – 31.270-901 – Belo Horizonte – Brasil.

katia.cardoso@hemominas.mg.gov.br, mba@eci.ufmg.br, vivianenpo@yahoo.com.br

Abstract. *Acquiring knowledge from experts has been a challenge within several research fields. There is no difference in the scope of Biomedicine, mainly because of the large amount of data. In this paper, we present an underway research being conducted in the domain of hematology, which investigates problems of the knowledge acquisition activity. A list of topics to carry out the knowledge acquisition was developed, which involve steps being applied in real world situations. The record of problems faced along the process aim to reach improvements in the best practices of the knowledge acquisition activity. In addition, we present and discuss partial results.*

Resumo. *Obter conhecimento especializado a partir de especialistas tem sido um desafio em diversos campos de pesquisa. Isso é especialmente verdadeiro em Biomedicina em função do volume de dados produzidos. No presente artigo, descreve-se pesquisa em andamento sobre dificuldades para aquisição do conhecimento no domínio da hematologia. Um roteiro para aquisição de conhecimento foi desenvolvido e tem sido aplicado empiricamente. Os registros das dificuldades ao longo do processo objetivam propor melhores práticas para a atividade. Resultados parciais são relatados e discutidos.*

1. Introdução

Investigar como o conhecimento produzido é traduzido em conhecimento de um campo científico, bem como os meios para organizá-lo e representá-lo é atividade essencial no âmbito da pesquisa. Entretanto, à medida que o volume de informação tem aumentado significativamente nos últimos 30 anos, representá-lo para uso por pessoas e por sistemas tem se tornando tarefa complexa.

Nesse contexto, ontologias têm sido propostas como alternativa para criação de representações da realidade. Desde os anos de 1980 têm sido estudadas como forma de representar conhecimento [Guarino, 1998], como referência para a criação de modelos [Fonseca, 2007] e para a representação do conhecimento [Vickery, 1997]. De forma simples, ontologias consistem de termos, relações e regras que regulam a combinação entre os termos organizados em uma taxonomia.

Ontologias são também definidas a partir do conjunto de processos que compõem a respectiva atividade de desenvolvimento [Fernandez, Gomez-Perez and Juristo, 1997],

como por exemplo, a aquisição de conhecimento. Trata-se de uma atividade sabidamente complexa em função das dificuldades de comunicação entre o engenheiro do conhecimento e o especialista [Milton et al 2006].

O presente trabalho se insere nesse contexto, enfocando a atividade de Aquisição de Conhecimento (AC) a partir de especialistas para o desenvolvimento de ontologias, conduzida no escopo de projeto em andamento no âmbito da biomedicina. Descreve-se a pesquisa que busca melhores práticas em atividades de AC, a partir da observação de dificuldades na obtenção de conhecimento de pesquisadores na Hemominas, instituição responsável pelo sistema hematológico e hemoterápico de Minas Gerais. A pesquisa busca, em última instância, estudar soluções para problemas como: a distância entre o conhecimento do especialista e o responsável pela AC - daqui em diante, “engenheiro” e barreiras entre especialista e engenheiro, como dificuldades de se obter consenso. O presente artigo apresenta o estágio atual da pesquisa: a metodologia utilizada para AC e as observações realizadas ao longo do processo na busca por respostas às perguntas citadas.

O restante do presente trabalho está organizado como segue: a seção 2 apresenta uma visão sobre AC. Cabe destacar que, por limitações de espaço, muitas questões são apenas pontuadas. A seção 3 apresenta a metodologia de pesquisa e a seção 4 discute resultados parciais. Finalmente, a seção 5 traz considerações e perspectivas futuras.

2. Aquisição de Conhecimento

A expressão AC é empregada desde a década de 1980 para se referir ao estudo da obtenção de *expertise* para representação em sistemas especialistas [Boose and Gaines, 1989] [Milton *et al*, 2006]. Diversas definições para AC são encontradas, mas é consenso que a atividade inclui pelo menos etapas de coleta, análise, estruturação e validação do conhecimento com finalidade de representação [Shadbolt and Burton, 1990].

A AC compreende um conjunto de tarefas que empregam teorias e métodos provenientes de campos diversos. Dentre essas disciplinas, cabe destacar a Ciência da Computação [Newell and Simon, 1975]; a Ciência Cognitiva [Hawkins, 1983]; a Linguística [Zellig Harris, 1976]; a Semiótica [Campbell, 1998] e a Psicologia [Kelly, 1955]. Cada um desses campos tem contribuído para a compreensão da atividade de AC.

2.1. Aquisição de Conhecimento: uma proposta de classificação das técnicas

Não há consenso sobre a classificação das técnicas de AC e diferentes propostas são encontradas na literatura, a saber: i) sob o ponto de vista do instrumento de aplicação: manual ou baseadas em computador (automáticas e interativas) [Boose and Gaines, 1989]; ii) em relação ao tipo de conhecimento obtido: procedural, conceitual, explícito, [Shadbolt, 2005] [Milton et al., 2006]; iii) do ponto de vista dos métodos [Shadbolt and Swallow, 1993]; e iv) dos métodos aplicados à biomedicina [Payne et al, 2007], esses últimos, relevantes no contexto desse trabalho.

As técnicas manuais, como a Grade de Repertórios, têm raiz na Psicologia [Boose and Gaines, 1989]. As técnicas interativas fazem uso de algum tipo de ferramenta para a interação do engenheiro com o especialista. Exemplos de técnicas automatizadas, como

o aprendizado da máquina, são apresentados no Projeto Neon¹ [Maynard and Nioche, 2006] [Gomez-Perez, Erdmann and Greaves, 2007]

As técnicas para AC orientadas para os métodos de aplicação podem variar de acordo com o tipo de conhecimento que se objetiva elicitar. Exemplos de tais técnicas são: a) geração de protocolos, entrevistas e observação; b) *técnicas baseadas em matriz*; c) *técnicas de ordenamento e*; d) *técnicas de limitação e restrição de informação* [Shadbolt and Swallow, 1993].

2.2. Aquisição de conhecimento em Biomedicina

A organização de terminologias na área biomédica é um desafio constante, em função da amplitude do assunto e da multiplicidade de interpretações para os dados e suas formas de obtê-los [Smith, 2008]. A literatura AC em biomedicina apresenta diferentes propostas para representação do conhecimento médico e biológico.

Vita et al (2006) descrevem a curadoria, parte do processo de anotação automatizado, quase uma exigência face à quantidade de artigos científicos produzidos. As anotações geradas exigem análise e validação por especialistas. Um exemplo é a extração de dados imunológicos que exigem alto nível de especialização. Stehr et al (2010) descrevem curadoria em redes colaborativas. Hoehndorf et al (2009) também se valem de uma interface *wiki* baseada em ontologias para aquisição semi-automática de conhecimento.

3. Metodologia de pesquisa

Os entrevistados para AC são especialistas do grupo de pesquisa sobre HTLV (vírus linfotrópico de células T humanas) em Minas Gerais. Um roteiro experimental para AC foi proposto, como forma de realizar a pesquisa fim: identificar dificuldades ao longo da atividade de AC em biomedicina. O roteiro foi elaborado a partir da revisão de literatura. Ao realizar uma AC real, o pesquisador vem registrando problemas de forma a propor melhorias. O roteiro de tarefas para AC contempla três fases: i) Levantamentos; ii) Contatos; 3) Validação.

A *fase de levantamentos* compreende atividades anteriores ao contato direto com especialistas. A primeira tarefa é conhecer o escopo da ontologia de domínio, conhecer a que fim a se destina, quem vai utilizá-la e com que propósitos, além de dados sobre a ontologia de alto nível. Além disso, é necessário identificar os especialistas, bem como o conhecimento que produzem e registram em fontes. Exemplos de levantamento dos especialistas e *expertise* são apresentados na Tabela 1.

Tabela 1. Extrato do levantamento de especialistas para atividade de AC

<i>Experts</i> ^(*)	A. B.C.F.	M.A R.	D.U.G.	E.F.B.	M.S.N.
Formação	Médico	Médico	Médico	Veterinário	Médico
Atuação	Pesquisa	Pesquisa e Medicina	Pesquisa e Ensino	Pesquisa e ensino	Pesquisa
Linhas de Pesquisa	Epidemiologia	Epidemiologia	Otoneurologia	-	Clinica médica
	Hematologia	Infectologia	Infectologia	-	Hematologia
	Virologia	-	-	Virologia	-

^(*) os especialistas são aqui identificados por códigos

¹ Disponível na internet em: <http://www.neon-project.org/> Acesso: 21 julho de 2011

A *fase de contatos* consiste no encontro com especialistas e realização de atividades que permitam obter e registrar conhecimento. Exemplos de técnicas utilizadas são: i) entrevistas, baseadas em *template* no *Protegé-Frames* baseado em Scheuermann et al (2009); ii) técnicas baseadas em matriz; iii) técnica de ordenamento,. A *fase de validação* faz uso de ferramentas *wiki* para colaboração na análise de termos candidatos a ontologia, bem como de suas definições. A partir do conhecimento obtido na AC, termos candidatos a ontologia são transpostos para uma *wiki*, onde em seguida são validados pelos especialistas via internet.

Tabela 2. roteiro de AC utilizado na pesquisa

Fase	Objetivo da tarefa	Descrição da Tarefa	Instrumento
(1) Levantamento	1.1 conhecer o contexto	Conhecer escopo da ontologia em desenvolvimento	Dados do projeto
	1.2 conhecer fundamentos	Conhecer conceitos básicos do domínio em questão	Literatura básica da área
	1.3 identificar <i>expertise</i>	Identificar <i>expertise</i> dos especialistas envolvidos	Diretórios de pesquisadores
(2) Contato	2.1 obter conhecimento	Entrevistas realizadas com especialistas	<i>Template Protege-Frames</i>
	2.2 conhecer terminologia	Identificar problemas para organização da informação	Técnicas de Matriz
	2.3 ver organização <i>ad-hoc</i>	Entender como especialistas ordenam conceitos	Técnicas de ordenamento
(3) Validação	3.1 validar conhecimento	Obter aprovação sobre termos adquiridos e suas definições	Página Wiki
	3.2 manter conhecimento	Atualizar dados após cada validação	Página Wiki

4. Resultados parciais

Conforme já mencionado, a pesquisa tem sido realizada no contexto de projeto de organização da informação sobre biomedicina. Ao longo da atividade de AC têm sido observadas e registradas problemas relativos aos especialistas e aos engenheiros. As principais observações são descritas brevemente à seguir. Os resultados parciais foram obtidos através de observação ao processo de AC em questão e, no atual estágio de pesquisa, não representam amostra quantitativa significativa.

Especialistas:

- O especialista tem pouco tempo disponível por acumular outras atividades como: ensino, coordenação, atendimento clínico entre outras;
- O especialista tem dificuldade de explicitar o que sabe e desconhece ontologias;
- Existe super-especialização no âmbito da própria especialidade, o que resulta em narrativas com diferentes níveis de granularidade e formas de organizar a informação;
- O especialista sugere fontes adicionais de conhecimento como complemento ao seu relato na AC, tendo em vista que suas inúmeras publicações;
- O uso de entidades provenientes de ontologias de alto nível, de forma a apresentar a organização hierárquica preliminar, acaba por confundir o especialista;

Engenheiro

- A curva de aprendizado no domínio da biomedicina é longa e complexa;
- O vocabulário especializado dificulta a compreensão das terminologias;
- O assunto é multidisciplinar e os especialistas atuam em sub-domínios: existem poucos especialistas para consulta e eles são super-especializados;
- Entidades de proveniências do alto nível usadas para orientar o engenheiro, nem sempre encontram respaldo no dia a dia do especialista, o que dificulta o diálogo.

As observações registradas até o momento resultam em recomendações: a) o uso de ferramenta de apoio como *Protegé-Frames* para organizar as entrevistas de acordo com princípios ontológicos sistemáticos; b) a consideração de AC semi-automática pelo grande volume de publicações; c) a imersão da literatura é acompanhada de técnicas como análise de assunto, análise conceitual, dentre outras; d) a realização da AC tem lugar ao longo das atividades rotineiras do especialista; e) uso de ferramentas interativas para que o especialista registre, ele mesmo, o que sabe, sem intervenção do engenheiro.

5. Considerações finais

Este artigo apresentou pesquisa em andamento que busca por melhorias nas práticas de AC. Para tal, descreveu-se literatura sobre AC (inclusive no âmbito da biomedicina), apresentou-se um roteiro para a AC, e discutiu-se resultados parciais através da identificação de dificuldades na comunicação.

A continuidade da pesquisa abordará questões ainda em aberto e buscará a confirmação de resultados parciais, na busca por respostas à questões mencionadas na introdução do presente artigo, a saber: como lidar com barreiras no processo de AC como: a formação de consenso; o desconhecimento sobre o trabalho do engenheiro; a falta de tempo do especialista? Que favorecem o consenso em ambientes colaborativos? Como lidar com diferenças terminológicas entre os próprios especialistas em ambientes colaborativos?

Agradecimentos

Este trabalho conta com apoio da Fundação Hemominas–e da FAPEMIG

Referências

- Boose, J. H., and Gaines, B. R. (1989) “Knowledge Acquisition for Knowledge-Based Systems: Notes on the State-of-the-Art”. <http://www.springerlink.com/>
- Campbell, K. E. et al. (1998) “Representing thoughts, words, and things in the UMLS”. <http://www.ncbi.nlm.nih.gov/pubmed/9760390>, October.
- Fernandez, M., Gomez-Perez, A. and Juristo, H. (1997) “Methontology: From Ontological Art Towards Ontological Engineering”, <http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-005.pdf>
- Fonseca, F. (2007) “The Double Role of Ontologies in Information Science Research”. <http://citeseerx.ist.psu.edu/>, April.
- Harris, Z. (1976) “On a theory of Language”. <http://www.jstor.org/stable/2025530>, May.

- Hawkins, D. (1983) “An analysis of expert thinking”. <http://www.sciencedirect.com/science/journal/00207373>, January.
- Hoehndorf, R. et al. (2009) “BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology”. <http://www.biomedcentral.com/1471-2105/10/S5/S5>, July.
- Hoffman, R.R. (1995) “Eliciting Knowledge from Experts: A methodological analysis”. *Organizational Behavioral and Human Decision Processes*. Vol. 62, no 2, p. 129-158.
- Gomez-Perez, J.M., Erdmann, M. and Greaves, M. (2007). “Applying problem solving methods for process knowledge acquisition, representation, and reasoning”. *Proceedings of the K-CAP 2007*
- Kelly, G. (1955). *Princípios da Psicologia dos Construtos Pessoais*. Norton, 3rd edition.
- Maynard, D. and Nioche, J. (2006) “Human Language Technology for Knowledge Acquisition for the Semantic Web”. *EKAW 2006*
- Milton, N., Clarke, D. and Shadbolt, N. (2006) “Knowledge engineering and psychology”. <http://portal.acm.org/citation.cfm?id=1221471>, December.
- Newell, A. and Simon, H.A. (1975) “Computer science as empirical inquiry: symbols and search” <http://citeseerx.ist.psu.edu/>, March.
- Nilsson, N. (2007) “The Physical Symbol System Hypothesis: Status and Prospects”. <http://ai.stanford.edu/~nilsson/OnlinePubs-Nils/PublishedPapers/pssh.pdf>
- Scheuermann, R. et al. (2009) “Toward an ontological treatment of disease and diagnosis”. http://ontology.buffalo.edu/medo/Disease_and_Diagnosis.pdf, February.
- Shadbolt, N. (2005) “Eliciting Expertise”. http://eprints.ecs.soton.ac.uk/14563/1/Elciting_Expertise.pdf,
- Smith, B. (2008). *New Desiderata for Biomedical Terminologies*. In *Applied Ontology*, pages 21–39. Ontos-Verlag.
- Stehr, H. et al. (2010) “PDBWiki: added value through community annotation of the Protein Data Bank”. <http://database.oxfordjournals.org/content/2010/baq009.full>, March.
- Vickery, B. C. (1986) “Knowledge representation: a brief review”. <http://www.emeraldinsight.com/>, September.
- Vita, R et al. (2006) “Curation of complex, context-dependent immunological data”. <http://www.biomedcentral.com/1471-2105/7/341>, July.
- Wolf, R. and Delugach, H. S. (1996) “Knowledge Acquisition via tracked repertory grids”. <http://www.cs.uah.edu/tech-reports/TR-UAH-CS-1996-02.pdf>