

Elicitação e Estruturação de Conhecimento: Um Arcabouço Conceitual Aplicado à Biodiversidade

Andréa C. F. Albuquerque^{1,2}, José L. Campos Dos Santos², Alberto N. Castro Jr¹

¹ Instituto de Ciência da Computação (IComp), Universidade Federal do Amazonas (UFAM), Manaus-AM, Brasil.

² Laboratório de Interoperabilidade Semântica (LIS), Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus-AM, Brasil

andreaalb.1993@gmail.com, lcampos@inpa.gov.br,
alberto@icompufam.edu.br

Abstract. *The knowledge (implicit-explicit), when modelled and made available on the Web, becomes essential for the generation of new knowledge. This work proposes a tacit knowledge formalization method to add semantics and expressiveness in formal ontologies to support the generation of scientific knowledge.*

Resumo. *O conhecimento, (implícito-explícito) quando modelado e disponibilizado na Web, torna-se essencial para a geração de novos conhecimentos. Este trabalho propõe um método de formalização de conhecimento tácito para agregar semântica e expressividade em ontologias formais no apoio a geração de conhecimento científico.*

1. Introdução

A recente abundância de informações, impõe o desafio de integrar domínios científicos e acelerar a busca de novas soluções para problemas de interesse e relevância social e de convergência ao mercado.

A biodiversidade, um campo de estudo transdisciplinar, propicia a ocorrência de heterogeneidade semântica, presente no problema de comunicação diária entre seres humanos, e sua presença nas ciências compromete seus resultados. Soluções para interoperabilidade são necessidades reais para pesquisa nesse domínio. Como exemplo, as classificações taxonômicas de espécies são alvo de discussão entre pesquisadores. Algumas árvores taxonômicas são claramente especificadas como em zoologia, para mamíferos. Entretanto, existem divergências de autores em vários domínios além de grande número de espécies para serem classificadas ou reclassificadas.

A tentativa de automatizar uma solução para as questões de heterogeneidade semântica é um grande desafio. Uma estratégia para lidar com tais problemas é o uso/reuso de ontologias para ajudar a eliminar conflitos semânticos e promover a integração. Ontologia define regras que formalmente descrevem o campo de interesse. Os dados podem ser quaisquer dados relacionados a esta área de interesse, que é extraído de várias fontes como base de dados, coleções de dados digitais, a Web etc. O mecanismo de inferência implanta regras em forma de axiomas, restrições, consequências lógicas e outros métodos baseados na definição formal da ontologia

sobre os dados reais para produzir mais informação do que já existe. A ontologia fornece ao engenheiro do conhecimento um vocabulário para expressar o domínio, através de seus termos, e um núcleo de conhecimento, fornecido por seus axiomas.

De modo geral, a ontologia pode estruturar o processo de aquisição de conhecimento para o propósito de compreensão, entendimento via processos automáticos e extração de conhecimento do ambiente da Web. Essas tecnologias, aplicadas ao domínio da biodiversidade, são um recurso valioso para o planejamento estratégico com impacto positivo, principalmente para a região Amazônica (i. e. avaliação de impacto ambiental, definição de áreas de preservação ambiental, proteção de espécies ameaçadas, recuperação de áreas degradadas, bioprospecção, estabelecimento de políticas públicas, legislação ambiental, entre outras).

Neste contexto, a OntoBio, uma ontologia formal de biodiversidade desenvolvida no cenário de coleções biológicas e eventos de coleta de campo do INPA, é uma ferramenta com funcionalidades de integração e estruturação de biodados, será utilizada como caso de estudo na viabilidade de se modelar o conhecimento tácito do especialista e incorporá-lo a um instrumento estruturante do conhecimento, tornando o mesmo mais expressivo semanticamente [Albuquerque, 2011; Santos *et al*, 2011].

2. Fundamentação Teórica

Este trabalho está baseado em três áreas de pesquisa interdependentes, que são: a) gestão do conhecimento – no trabalho de T. D. Wilson sobre a natureza da gestão do conhecimento [Tarapanoff, 2006, p.37]; b) o processo de aquisição do conhecimento – na pesquisa de E. A. D. Moresi sobre o mapeamento do conhecimento para geração de novo conhecimento [Tarapanoff, 2006, p.277]; e c) técnicas de elicitação de conhecimento – baseado nos estudos de J. F. A. Gaspar e R. C. R. Miranda que trata da captura de conhecimento de fontes distintas [Tarapanoff, 2006, p.391].

3. Arcabouço Conceitual para Elicitação e Estruturação do Conhecimento

O sucesso de uma integração semântica de dados é dependente do poder de expressividade da(s) ontologia(s) utilizada(s). No entanto, a escassez de ontologias de domínio sofisticadas, inviabiliza a integração efetiva dos dados. O arcabouço conceitual apresentado na Figura 1, foi concebido com o objetivo de elicitar conhecimento tácito, destinado a agregar expressividade semântica à ontologias formais e bases de dados científicas, considerando-as como um dos elementos estruturantes para integração e organização de diferentes modelos mentais do especialista, possibilitando o processamento do conhecimento e sua gestão, aplicado a dados de biodiversidade. O propósito é fornecer um ferramental tecnológico para facilitar e enriquecer o processo de elicitação do conhecimento que servirá de parâmetro de entrada para o desenvolvimento de ontologias mais expressivas e ricas semanticamente.

Um fato que chama atenção neste universo de dados sobre biodiversidade disponíveis na Web, é a inexistência de dados não-estruturados como o conhecimento legado, (ex.: do pesquisador, do mateiro, do pescador, do guia nativo, entre outros), potenciais para agregar valor às bases disponíveis. Estes dados encontram-se no modelo mental do indivíduo e tendem a se perder diante da entropia da informação. Para ilustrar este contexto, pode-se apresentar, por exemplo, um levantamento de requisitos com um

especialista de *aracnídeos* (aranhas), que relata que sempre há ocorrência de aranhas em regiões de várzea na Amazônia em raízes submersas, isto baseado em sua experiência de especialista. Este dado não está expresso (e nem poderia ainda) na OntoBio, por se tratar de conhecimento adquirido com os anos de pesquisa e experiência do especialista (tácito), mas é informação relevante que poderá orientar uma série de ações futuras.

Os benefícios de uma integração de dados científicos mais expressivos, permitirão a biociência, avançar rapidamente em áreas até hoje não investigadas. Por esta razão, faz-se necessário integrar o modelo mental do especialista às bases de dados de biodiversidade existentes e as já integradas na OntoBio. Atualmente, a OntoBio integra apenas as bases de dados estruturadas de biodiversidade, não considerando o conhecimento tácito de especialistas.

Um arcabouço conceitual orientará a elicitación e integração do conhecimento tácito, permitindo o desenvolvimento de ontologias mais ricas e expressivas semanticamente, tornando o processo de aquisição e uso de conhecimento para múltiplos propósitos. Ainda, permitirá agregar à OntoBio, além de sintaxe, estruturação e semântica, um modelo de estrutura conceitual e cognitiva do especialista do domínio. O arcabouço deve contemplar modelos mentais multidisciplinares e por isso se faz necessário a extração de modelos conceituais e cognitivos para um domínio compatível com a OntoBio. O resultado extraído poderá ser mapeado para a ontologia desejada afim de ser validado. Na Figura 1 é apresentado o método proposto para elicitación do conhecimento que compreende 5 etapas:

- (E₁) O arcabouço deverá elicitar os Modelos Mentais dos Especialistas (MMEs);
- (E₂) Os MMEs deverão ser modelados conceitualmente através de um processo de formalização, gerando as Bases de Conhecimento Especialista (BCEs);
- (E₃) Mecanismo de composição das BCEs, que com o uso de Ferramentas Inteligentes (FIs), poderá desempenhar função de comunicação, negociação, manutenção da verdade e integração, além de considerar a natureza das BCEs (independentes, conflitantes, complementares, etc.);
- (E₄) Especificar conjunto de Esquemas de Conceitualização bem Fundamentada (ECFs) das BCEs, permitindo diretivas, acopláveis à estrutura de ontologias formais e às bases de dados não convencionais de biodiversidade;
- (E₅) Validação do arcabouço conceitual através da integração dos ECFs às bases de dados não-convencionais e à estrutura formal da OntoBio, incrementando o processo de aquisição do conhecimento, incorporando sintaxe, estrutura, semântica e disponibilizando elementos para anotação semântica de dados e estruturas que podem ser vinculadas à ontologias abertas que poderão auxiliar a publicação dessas informações na *Linked Open Data* (LOD).

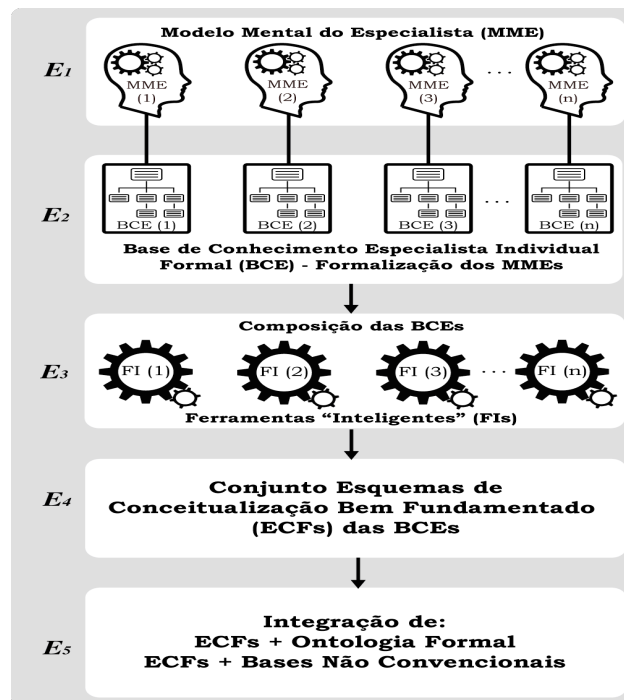


Figura 1 – Arcabouço Conceitual.

4. Considerações Finais

A modelagem de conhecimento tácito continua sendo uma área com muitas questões não respondidas e por isso solicita ainda muita pesquisa. No contexto de biodiversidade, a formalização do conhecimento legado do especialista, assim como sua incorporação às bases de dados existentes através de instrumentos estruturantes de conhecimento, resultaria numa expansão do conjunto de possibilidades de pesquisa. Essa abordagem evidencia uma evolução da disseminação de dados, baseada em padrões de dados e metadados com limitado poder de expressividade, que não são interoperáveis.

5. Agradecimentos

Ao IComp/UFAM, LIS/INPA, FAPEAM (Fundação de Amparo à Pesquisa do Estado do Amazonas) - concessão 021/2011 062.03101/2012-DO, e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) - concessão 486333/2011-6 por financiarem parcialmente esta pesquisa.

Referências

- Albuquerque, A.C.F. (2011) “Desenvolvimento de uma Ontologia de Domínio para Modelagem de Biodiversidade“. Dissertação de Mestrado, UFAM.
- Tarapanoff, K. (2006) “Inteligência, Informação e Conhecimento em Corporações“. Brasília: IBICT, UNESCO. ISBN: 85-7652-063-x.
- Santos, J.L.C. *et al* (2011) “Ontologias para Interoperabilidade de Modelos e Sistemas de Informação de Biodiversidade“. In: Proceedings of the Iberoamerican Meeting of Ontological Research collocated CITA 2011. Aachen: CEURWS.org. v.728. Gramado.