

Ontologias, Web semântica e saúde

DOI: 10.3395/reciis.v3i1.238pt



Frederico Freitas

Centro de Informática,
Universidade Federal de
Pernambuco, Recife, Brasil
fred@cin.ufpe.br



Stefan Schulz

Instituto de Biometria
Médica e Informática da
Medicina, Centro Médico
Universitário de Freiburg,
Freiburg, Alemanha
stschulz@uni-freiburg.de

Com uma longa tradição de estudo no ramo da filosofia, que remonta há pelo menos remotos 23 séculos, o termo “Ontologia” transformou-se num dos termos mais em moda no mundo da informática de hoje, sendo aplicado em sistemas de muitos outros campos, incluindo-se, entre os principais, biologia e medicina. Enquanto o termo “Ontologia” outrora denotava um ramo da metafísica, “ontologias” são hoje entendidas como vocabulários formais que descrevem as premissas básicas de um determinado domínio.

Há pelo menos uma razão principal para todo este interesse por parte dos informatas. Segundo Tim Berners-Lee, um dos principais responsáveis pela sedimentação da própria internet, as ontologias constituem o componente e a motivação principais da Web semântica, uma Web em que os programas e sistemas são capazes de “entender” e processar dados das páginas, de acordo com o contexto. Mas o que são ontologias e como elas ajudam os sistemas a conseguir processar os dados com tal profundidade?

Em palavras práticas, ontologias padronizam significado através de identificadores semânticos, os quais podem representar o mundo real e conceitual. Ontologias constituem-se em definições de conceitos, classes, propriedades, relações, restrições e axiomas sobre um determinado domínio (por exemplo, neurologia).

Ao longo dos artigos deste número temático são apresentados vários exemplos de ontologias, que variam em complexidade, expressividade etc. Então, se uma

página da Web referencia uma ou mais ontologias (por exemplo, dizendo que um determinado trecho da página é um nome de professor, referenciando este conceito de uma ontologia sobre Academia), ela está atribuindo significado ao conteúdo da página de forma que um software poderá usar as outras definições da ontologia que definem as relações e restrições *a priori* relacionadas ao conceito referenciado e interpretará outras partes da página dentro do contexto definido pela ontologia. Por exemplo, ao achar uma lista de alunos de doutorado na página do professor, o software poderá interpretar que aqueles alunos são orientados pelo professor, que trabalham em áreas que o professor pesquisa (e que podem também estar na página do professor), e que estão matriculados no programa de pós-graduação ao qual o professor pertence. Digno de nota é o fato de que estas informações podem não estar literalmente presentes na página Web. Portanto, processar informações usando ontologias, que provêem excelente contexto para o entendimento das informações, tanto para usuários humanos quanto para agentes de software, vem se tornando uma tendência em várias áreas e tipos de aplicações. A Web semântica, por exemplo, deve ter um forte impacto sobre os sítios comerciais da Web; ontologias também servem como vocabulário em uma troca de mensagens entre os ditos agentes inteligentes, entidades de software que podem raciocinar sobre conhecimento, permitindo que estes agentes possam negociar pedidos de compras em nome de suas empresas – e já existem protótipos acadêmicos simulando esta situação.

Entre as áreas mais comuns de aplicação de ontologia estão medicina e biologia, existindo, inclusive, alguns centros quase que exclusivamente dedicados ao estudo desta tecnologia. Nos Estados Unidos, o Departamento de Informática Médica da Universidade de Stanford, criou o mais empregado editor de ontologias, o Protégé. Existe também uma organização virtual, o Centro Nacional de Pesquisas em Ontologias Biomédicas (National Center for Biomedical Ontologies), que envolve os grupos de pesquisa das universidades de Stanford, Victoria e Buffalo, além da Mayo Clinic, em Rochester. Na Europa, existem o Grupo de Pesquisa em Ontologias em Medicina e Ciências Biológicas no Instituto de Informática Médica, Estatística e Epidemiologia, em Leipzig, e o Grupo de Pesquisa em Informática Médica no Centro Médico Universitário de Freiburg, ambos na Alemanha, e o Grupo em Informática Biológica e Saúde, na Universidade de Manchester, na Inglaterra, estão entre os que mais pesquisam em ontologias biológicas e Web Semântica. A complexidade do conhecimento médico e biológico torna difícil a confecção de sistemas tradicionais, pois para assistir tarefas médicas, os sistemas precisam de muito conhecimento e capacidade de inferência. Esta talvez seja a principal razão da aplicabilidade e conseqüente sucesso do uso de ontologias nestas áreas. Discorremos um pouco mais acerca destas necessidades no primeiro artigo. Apenas cabe ressaltar que o volume de pesquisas produzido ligando ontologias, Web semântica e saúde justificaram o lançamento deste número temático e aceitamos a tarefa, convencidos de que a RECIIS é, de fato, um veículo em que todos estes tópicos de pesquisa são contemplados. A comunidade que pesquisa nestes tópicos respondeu à chamada de trabalhos satisfatoriamente e os editores se sentem agradecidos com as contribuições que nos foram enviadas.

Das submissões recebidas para este número temático, foram selecionados cinco artigos pelos revisores, que, juntados a este artigo introdutório dos editores convidados, compõem os seis artigos deste número temático. Durante a confecção do número temático, procuramos ordenar os artigos de forma a ir apresentando gradativamente os conceitos e práticas em ontologias biomédicas em grau crescente de complexidade, uma vez que os dois primeiros temas do número temático – Ontologias e Web semântica - ainda não são muito difundidos no Brasil. Abaixo, descreveremos brevemente cada um dos artigos do número temático.

O primeiro artigo, intitulado “Levantamento das atuais terminologias e ontologias em biologia e medicina”, tem por objetivo apresentar conceitos introdutórios sobre ontologias e sistemas terminológicos em biologia e medicina, proporcionando uma visão abrangente do assunto pela descrição de sistemas mais difundidos e/ou emblemáticos. Após uma breve discussão sobre o uso de terminologias em contraste com o uso de ontologias, são apresentados os sistemas CID (Classificação Internacional de Doenças), SNOMED CT (Nomenclatura Sistematizada de Termos médico-clínicos), MeSH (Descritores de Saúde), openGalen (Arquitetura Geral para Linguagens, Enciclopédias e Nomenclaturas in

Medicine, FMA (Modelo Fundacional de Anatomia), bem como as iniciativas UMLS (Sistema Unificado de Linguagem Médica), a OBO Foundry (Fundação de Ontologias Biomédicas Aberta”) em geral e a Ontologia Gênica (GO) em particular.

O artigo “Bases ontológicas e conceituais para um modelo do conhecimento científico em artigos biomédicos”, escrito por Carlos Marcondes e co-autores, propõe uma classificação de artigos científicos da área de Medicina. Cada classe possui um modelo de anotação semântica, anotações estas que devem ser efetuadas usando ontologias, de forma que o conhecimento expresso nos artigos consiga ser processado e “entendido” por sistemas computacionais. Antes da proposição, o artigo traz uma discussão que justifica a existência do modelo a partir de teorias de Metodologia Científica e Filosofia da Ciência e da análise de uma base de 75 artigos médicos. O modelo habilita a recuperação semântica de informações de artigos, permitindo buscá-lo através de conceitos e relações, ao invés de palavras-chave como acontece nos engenhos de busca tradicionais.

O artigo “Vantagens e limitações de ontologias formais no domínio biomédico”, elaborado por Stefan Schulz e co-autores, dentre os quais Barry Smith, filósofo e autor de vários livros e artigos bastante conhecidos sobre bio-ontologias e diretor de várias organizações relacionadas ao assunto, como a própria Fundação OBO, serve como excelente introdução para os formalismos matemáticos usados em ontologias biomédicas, propondo uma série de critérios para delimitar o conceito de ontologia de recursos de representação de conhecimento de uma forma mais abrangente. As vantagens e desvantagens de cada representação são cuidadosamente discutidas, a partir de tesouros, fonte de conhecimento freqüentemente disponível na área biomédica, até a expressiva lógica de descrições - o mais expressivo formalismo de representação de conhecimento padronizado pelo Consórcio da Web (W3C). A discussão gira, primordialmente, em torno da relação custo-benefício em adicionar-se expressividade à forma de representação em uso. Exemplos simples usando lógica de descrição são introduzidos, em complexidade gradativa, desta forma ajudando o leitor a ter uma idéia concreta de como são representados conceitos biomédicos complexos e de como estes conceitos são usados durante o processo de raciocínio automático, bem como possíveis resvalos de modelagem que podem levar a raciocínios automáticos incorretos.


O artigo seguinte, “Uma análise ontológica do eletrocardiograma” por Bernardo Gonçalves e co-autores, traz a descrição de uma ontologia cuidadosamente construída para possibilitar o raciocínio sobre um domínio complexo, específico e desafiador em termos de representação, a execução de um eletrocardiograma. A ontologia foi elaborada empregando-se duas ontologias biomédicas bastante ricas e populares, a OBO e a FMA - já descritas no primeiro artigo - e a ontologia de topo UFO (Ontologia Fundacional Unificada). Esta ontologia de topo empresta bases sólidas para a representação de elementos complexos presentes direta ou indiretamente

em eletrocardiogramas, tais como relações parte-todo universais e relações temporais de produção, caracterização e geração, classificando-as com suas meta-características adequadas, sejam elas simétricas ou anti-simétricas, reflexivas ou irreflexivas e transitivas ou intransitivas. Portanto, a elaboração da ontologia ECG constitui um bom exemplo de como empregar princípios de modelagem bem fundamentados para garantir que o raciocinador automático não derive conclusões errôneas por falta de conhecimentos básicos, como as conseqüências de um conceito ser parte de outro. Ao fim do artigo, são sugeridas algumas formas de aplicação do conhecimento disponibilizado.

Os últimos dois artigos são dedicados à construção de ontologias, mas não sobre o ponto de modelagem. O artigo “Aspectos metodológicos no reuso de ontologias: um estudo a partir das anotações genômicas no domínio dos tripanosomatídeos”, de Maria Luiza Campos e co-autores, discute em linhas gerais o reuso de ontologias, descrevendo ainda seus experimentos na área biomédica, a descrição do genoma de tripanosomatídeos. Foi empregado um alinhador, cuja função é comparar duas ontologias, mostrando as correspondências entre os elementos destas ontologias, uma tarefa correlata à tarefa de reuso. Os autores findam o artigo com uma lista de conclusões acerca de seu experimento específico, ou seja, em que casos o alinhador correspondeu às expectativas dos autores.

Finalmente, o artigo “Extração automática de termos compostos para construção de ontologias: um experimento na área da saúde”, confeccionado por Lucelene Lopes e Renata Vieira, discorre sobre uma ferramenta de aprendizado automático de ontologias,

elaborada dentro do grupo de pesquisa dos autores. A ferramenta está fundamentada em técnicas específicas de uma subárea específica de estudo de Inteligência Artificial, conhecida como Processamento de Linguagem Natural, que se debruça sobre o processamento de textos do ponto de vista lingüístico, acarretando, portanto, em um processamento muito mais profundo das informações textuais. Estas técnicas são usadas para extrair elementos que irão compor uma ontologia. Este tipo de abordagem é particularmente interessante por automatizar o processo de produção de conhecimento para as ontologias, acelerando o processo de obtê-las. As autoras descrevem experimentos iniciais com *corpus* sobre pediatria. Os experimentos mostraram que de fato abordagens baseadas em aprendizado podem ser úteis na etapa de construção da terminologia para uma ontologia, não garantindo, porém, a cobertura necessária para incluir toda a terminologia contida no *corpus*.

Como últimas palavras, gostaríamos de agradecer aos revisores, Guilherme Ataíde, Werner Ceusters, Ronald Cornet, Marcos Galindo, Rosario Girardi, Giancarlo Guizzardi, Robert Hoehndorf e César Tacla, por realizarem um trabalho sério e de boa qualidade, sem o qual a produção deste número temático provavelmente não seria possível. Os editores convidados dedicam ainda um agradecimento aos editores científicos da RECIIS, Carlos Saldanha Machado e Josué Laguardia, por nos guiarem durante todo o processo de produção da edição, com várias dicas úteis, sendo muito atenciosos e respondendo quase imediatamente a uma longa lista de dúvidas nossas surgidas ao longo do trabalho, além de brindar-nos com a oportunidade de publicá-lo. 

Sobre os editores

Fred Freitas

É PhD pela Universidade de Santa Catarina, Brasil, e atualmente é afiliado ao Centro de Informática da Universidade Federal de Pernambuco, Brasil (CIn/UFPE). Conduziu pesquisas por quase um ano no Departamento de Informática da Universidade de Karlsruhe, como integrante do projeto Brasil-Alemanha “*A semantic approach to data retrieval*” (Abordagem semântica da recuperação de dados). Publicou diversos artigos em conferências e seminários de renome, como IJCAI e outros patrocinados pela ACM (*Association on Computer Machinery*) e pelo IEEE (*Institute of Electrical and Electronical Engineering*). Co-presidiu duas séries de seminários: O WONTO (*Workshop on Ontologies and their Applications/Seminário de Ontologias e Suas Aplicações*), no Brasil, e o BAOSW (*Building Applications with Ontologies for the Semantic Web/Construção de Aplicações com Ontologias para a Semantic Web*), em Portugal. Co-editou Edições Especiais sobre temas relacionados do JBCS (*Journal of Brazilian Computer Society*) e do JUCS (*Journal of Universal Computer Science*). Colabora, atualmente, com a Universidade de Paul Cessane em Marselha, e INRIA, Montbonnot, na França, e as Universidades de Karlsruhe, Freiburg e Mannheim, na Alemanha. Suas áreas de interesse incluem ontologias, sistemas multiagentes, representação de conhecimento, mediação, e mineração de texto.

Stefan Schulz

É formado em medicina pela Heidelberg University, Alemanha, e é pesquisador sênior e professor do Instituto de Biometria Médica e Informática da Medicina do Centro Médico Universitário Freiburg, onde chefia o Grupo de Pesquisas em Informática na Medicina. Seu trabalho se concentra em terminologias e ontologias biomédicas, representação do conhecimento biomédico, recuperação de documentos médicos multilíngües, mineração de texto e dados em repositórios de documentos clínicos, aprendizado eletrônico na Medicina, e informática da saúde em países em desenvolvimento.

Após executar trabalhos clínicos em cirurgia e medicina interna, obteve seu diploma de doutorado na área da higiene tropical, onde efetuou um estudo de campo parasitológico em São Luís, Brasil. Após obter qualificação técnica em computação médica, mudou-se para a Universidade de Freiburg, onde participou de projetos de desenvolvimento de software clínico e educacional, e de diversos projetos de pesquisa na área da extração de informações, terminologias biomédicas, engenharia da linguagem médica, e tecnologias semânticas. Tem desempenhado papéis de liderança em diversos projetos financiados pela União Européia. Stefan Schulz é autor de mais de cem publicações revisadas por especialistas, e recebeu vários prêmios. Tem oferecido repetidas contribuições a projetos de pesquisa na área da informática de saúde brasileira desde 2001, como pesquisador convidado da Pontifícia Universidade Católica do Paraná (PUC-PR).