

Universidade Federal do Estado do Rio de Janeiro - UNIRIO

Programa de Pós-Graduação em Informática

Mestrado em Sistemas de Informação

Metodologia para Modelagem Conceitual de Dados

Lúcia Castro

(lucia.castro@uniriotec.br)

Orientadora: *Fernanda Baião (fernanda.baiao@uniriotec.br)*

Co-orientador: *Giancarlo Guizzardi (gguizzardi@gmail.com)*

Ingresso: *março de 2008*

Conclusão: *dezembro de 2009*

Resumo. *Modelos conceituais semanticamente ricos são a base para a integração semântica de dados. Embora os modelos conceituais existam há mais de 30 anos, pouco foi proposto em termos de um processo para a modelagem de dados. A construção de um modelo de dados implica na aquisição dos conceitos de um universo do discurso por parte do projetista, e sua posterior transcrição em uma linguagem de modelagem; para tanto, é necessário um método. A linguagem empregada deve ter construtos suficientes para a descrição semântica dos conceitos envolvidos. Uma análise da linguagem, a partir de princípios lingüísticos que se aplicam às linguagens naturais, pode atestar sua adequação à representação de modelos conceituais. Este trabalho propõe uma metodologia para a modelagem conceitual de dados que se baseia nos processos cognitivos humanos. A representação do modelo construído utiliza uma linguagem desenvolvida a partir de uma ontologia de fundamentação, que será, ao longo do trabalho, comparada ao conceito de gramática universal*

Palavras-chave. *Modelagem conceitual, cognição, ontologia, integração semântica de dados.*

1. Introdução

Em um mundo globalizado, a palavra de ordem é integração: a fusão de empresas, uma constante no mundo corporativo, ou mesmo a otimização de processos de negócios dentro de uma mesma empresa, implicam na integração de dados corporativos. Mais ainda, a Web Semântica é uma realidade mas, para que ela alcance seus objetivos, a integração é “o nome do jogo”, como diz Berners-Lee [2008]. Além das questões estruturais envolvidas na integração de dados (que estão fora do escopo deste trabalho), há que se considerar seus aspectos semânticos, o que envolve várias dificuldades. A comparação semântica de esquemas é uma tarefa manual ou, na melhor das hipóteses, semi-automática, visto que envolve consultas aos usuários e a dedução de equivalências a partir de antigas fontes de informação, nem sempre confiáveis [Kent 1998]. Conseqüentemente, a comparação de esquemas tende a ser uma tarefa lenta e dispendiosa, embora de vital importância no que diz respeito à integração semântica de dados – não há como afirmar a equivalência semântica entre conceitos a menos que se comparem esses conceitos, com base no que estiver documentado em esquemas e metadados.

Os modelos conceituais têm sido tratados desde que Chen [1990] descreveu a abordagem entidade-relacionamento para a sua construção. A literatura especializada define os modelos conceituais, em geral, como uma coleção de conceitos que podem ser usados para descrever um conjunto de dados e as operações usadas para manipular esses dados [Batini, Ceri e Navathe 1992]; ou uma ferramenta para representar a realidade [Elmasri e Navathe 2007]; ou ainda “*uma descrição do banco de dados de forma independente de implementação em um SGBD*” [Heuser 2001]. No entanto, a literatura aqui citada não trata de como esses conceitos, ou “esquema da empresa” [Chen 1990], devem ser transmitidos ao projetista responsável pela elaboração do modelo. Em alguns trabalhos [Elmasri e Navathe 2007], existe a alusão a uma entrevista a ser conduzida pelo projetista com o usuário conhecedor dos conceitos, ou esquema da empresa; entretanto, nada é dito sobre o processo de aquisição de conhecimento por parte do projetista, e a ênfase recai quase que exclusivamente sobre a descrição do modelo.

Para evitar problemas de integração, é necessário que se construam modelos conceituais semanticamente ricos que permitam uma comparação automática de seus conceitos. A criação de modelos conceituais com essas características implica na definição de um método para a aquisição do conhecimento do domínio modelado por parte do projetista, bem como sua representação através de uma linguagem ontológica. Kent [1998] afirma que o homem luta para impor aos computadores sua maneira de perceber e processar informações sobre o ambiente que o cerca. Ainda segundo ele, a dificuldade talvez seja que a informação, bem como a linguagem que os homens usam para representá-la, é, em sua essência, amorfa, ambígua e subjetiva demais para ser enquadrada nos processos determinísticos de um computador. Questões como ambigüidade e subjetividade são tratadas de maneira automática por humanos, mas o mesmo não ocorre quando o processamento de dados é feito por máquinas.

Este trabalho tem como objetivo a proposta de um processo para a modelagem conceitual. O trabalho de aquisição de conceitos por parte do projetista do modelo, através da entrevista com o usuário, tem como base o MEDS, Método de Explicitação

do Discurso Subjacente [Nicolaci-da-Costa 2007]. Já a verificação da adequação lingüística de uma linguagem para a representação dos conceitos toma por base a UFO, Unified Foundational Ontology [Guizzardi 2005]. A fundamentação teórica para o trabalho é apresentada nas seções de 2 a 5, que abordam os seguintes assuntos, respectivamente: modelos e modelagem, a linguagem e a cognição humanas, semiologia e lingüística, linguagem de representação e UFO. A seção 6 discute as contribuições previstas, o estado atual do trabalho e a metodologia a ser aplicada para a avaliação de resultados. A seção 7 conclui o artigo.

2. Modelos e Modelagem

Chen [1990] propõe o Método Entidade-Relacionamento para, segundo ele mesmo, “*acrescentar um estágio intermediário ao projeto lógico de bancos de dados.*” Ainda de acordo com ele, “*O projetista de banco de dados primeiro identifica as entidades e os relacionamentos que são de interesse para a empresa, usando a técnica diagramática Entidade-Relacionamento (E-R). [...] O esquema da empresa deve ser uma representação ‘pura’ do mundo real e deve ser independente de considerações sobre armazenamento e eficiência.*” Essa é uma descrição para o modelo conceitual de dados que não trata, porém, de como o projetista deve agir para conhecer as entidades e relacionamentos “que são de interesse da empresa”.

Se o projetista do modelo não tem, normalmente, conhecimento dos conceitos relativos ao domínio a ser modelado, e se o usuário, por sua vez, não tem conhecimento sobre modelagem, cria-se um impasse: a menos que a transmissão das informações ao projetista seja feita de maneira clara e eficiente, o modelo gerado pode não corresponder à realidade do usuário. Nesse processo, tanto o projetista quanto o usuário podem ser traídos por suas experiências anteriores. O projetista habituado a trabalhar com um determinado modelo tende a querer “encaixar” os domínios nesse modelo, ignorando, inadvertidamente, conceitos cujos construtos usados não são capazes de representar. O usuário, por sua vez, pode ser traído por seu excesso de conhecimento – por realizar suas tarefas de maneira automática, pode deixar de explicitar detalhes de execução que não mais percebe [Sternberg 2008]. Sem a adoção de uma técnica ou roteiro de trabalho que evite ou minimize as chances de erros por parte do projetista e do usuário, o risco de insucesso é razoável. Após a definição da web semântica, uma série de trabalhos tem proposto metodologias para a criação de ontologias [Noy e McGuinness, 2001, Fernández *et al.*, 1997, Orlean, 2003], no entanto nenhuma delas foca na utilização de uma linguagem de fundamentação para a representação do modelo descrito. Neste sentido, o presente trabalho pode ser adaptado para compor quaisquer das metodologias existentes de criação de ontologias.

Nicolaci-da-Costa [2007] descreve o MEDS – Método de Explicitação do Discurso Subjacente. Resultante da integração de procedimentos empregados na prática de pesquisas em psicologia clínica, e sob a influência de métodos oriundos da psicanálise, o MEDS adota a técnica da *livre escuta* em entrevistas para a coleta de dados no campo da pesquisa qualitativa. Inspirado também no princípio da *associação livre*, o MEDS parte da premissa de que o que é importante para uma pessoa sobre determinado assunto inevitavelmente surge no discurso espontâneo dessa pessoa. Por suas características, o MEDS pode ser usado como base para a elaboração das entrevistas com os usuários de um determinado domínio.

3. Linguagem e Cognição

Jackendoff [1983] afirma que estudar a semântica em uma linguagem natural é estudar a psicologia cognitiva, e que a estrutura gramatical da linguagem natural é uma fonte de evidências importante para a teoria da cognição. De acordo com Sternberg [2008], “A unidade fundamental de conhecimento simbólico é o conceito – uma idéia a respeito de algo que oferece um meio de entender o mundo”. O ser humano percebe seu ambiente a partir de seus sentidos, principalmente pela visão. O conhecimento simbólico humano é composto de conceitos, que os seres humanos organizam em categorias, que são definidas a partir de uma característica (ou conjunto de características), definitiva, ou seja, a característica que a determina. Estudos sobre as linguagens naturais e sobre a cognição humana são de fundamental importância para a metodologia de modelagem de dados; esse estudo será aplicado na avaliação da linguagem de representação a ser adotada na modelagem.

4. Semiologia e Lingüística

Saussure [2006], em sua enunciação da semiologia, ou semiótica, datada de 1915, afirma que “Ela nos ensinará em que consistem os signos, que leis os regem. [...] A Lingüística não é senão uma parte dessa ciência geral; as leis que a Semiologia descobrir serão aplicáveis à Lingüística e esta se achará dessarte vinculada a um domínio bem definido no conjunto dos fatos humanos. [...] Cabe ao psicólogo determinar o lugar exato da Semiologia; a tarefa do lingüista é definir o que faz da língua um sistema especial no conjunto dos fatos semiológicos.” A Semiologia engloba a semântica (estudo da relação entre os signos e seus significados), a sintaxe (estudo das relações formais ou estruturais entre os signos) e a pragmática (estudo das relações entre o signo e seus usuários); já a Lingüística ocupa-se do estudo do signo lingüístico. O filósofo Charles Peirce representa o signo através de uma tríade [Chandler 1995], que é comumente conhecida como Triângulo de Ullmann (referência ao lingüista húngaro Stephen Ullmann), ou Triângulo Semiótico (figura 1):

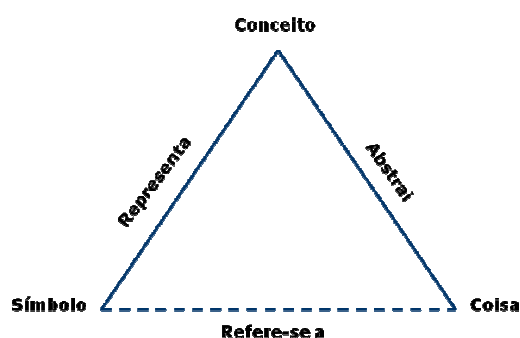


Figura 1. O Triângulo Semiótico

Um modelo conceitual semanticamente rico deve ser expresso em uma linguagem que possua os construtos (símbolos) necessários à representação de todos os conceitos integrantes do domínio modelado, visto que esses conceitos abstraem a realidade (coisa) sobre a qual se deseja armazenar informações; em outras palavras, para que um modelo conceitual seja semanticamente rico, a linguagem que o representa deve ser adequada.

5. Linguagem de Representação - UFO

Segundo Hitchman [2002], uma linguagem, que ele descreve como sendo um instrumento de medição que nos permite ver os componentes ocultos de uma estrutura conceitual, é o meio para a modelagem de dados. Sternberg [2008] afirma que uma linguagem deve ser *comunicativa, arbitrariamente simbólica, regularmente estruturada, estruturada em múltiplos níveis, gerativa, produtiva e dinâmica*. A linguagem de representação dos modelos conceituais deve possuir as mesmas propriedades das linguagens naturais em que os conceitos do ambiente a ser modelado são construídos, a fim de manter a representação o mais fidedigna possível.

Ontologias formais são uma estrutura de representação do conhecimento que despertam grande interesse em pesquisas na Ciência da Informação. Azevedo *et al* [2009] apresentam uma abordagem para gestão de ontologias, e ressaltam que as promessas de integração e de interoperabilidade ainda não foram plenamente alcançadas. Guizzardi [2005] apresenta uma ontologia de fundamentação para os conceitos da modelagem conceitual; essa ontologia – UFO, ou Unified Foundational Ontology – apresenta-se estratificada em UFO-A, UFO-B e UFO-C, sendo a primeira uma ontologia essencialmente de fundamentação, apresentando as características e a base para o desenvolvimento de ontologias de domínio, bem como os parâmetros para a sua avaliação. Por seu caráter de fundamentação, a UFO compreende uma riqueza de construtos que habilita seu uso na representação de qualquer domínio, permitindo, assim, a construção de modelos semanticamente ricos.

6. Contribuição e estado atual

O presente trabalho visa descrever uma metodologia para a modelagem conceitual de dados, incluindo técnicas para a aquisição de conceitos do domínio a ser modelado por parte do projetista, o método proposto e a análise de uma linguagem de fundamentação para a representação do modelo descrito. A defesa da proposta foi concluída recentemente. O trabalho será concluído com um estudo de caso a ser realizado no IBGE – Instituto Brasileiro de Geografia e Estatística. Tal estudo constituirá da aplicação do método e da linguagem de representação por parte de projetistas da instituição, visando a geração de modelos e de metadados semanticamente ricos para as bases de dados operacionais, e da posterior comparação dos modelos produzidos com os existentes. Com isso, busca-se investigar o ganho de semântica. A figura 2 apresenta o gráfico da proposta geral de estudo.

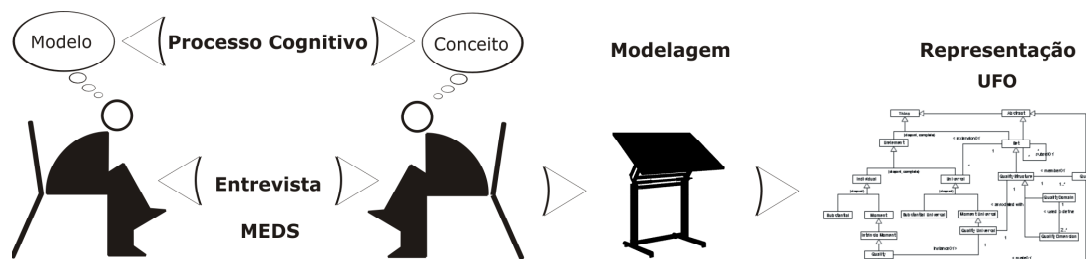


Figura 2. Gráfico da Proposta

7. Conclusão

O presente trabalho tem por foco o desenvolvimento de uma metodologia para a modelagem conceitual de dados, com vistas a criar modelos mais ricos, e evitar problemas semânticos quando da integração de dados. Para tanto, o trabalho parte de

uma análise de como os dados e seus referentes no “mundo real” podem ser como os falsos cognatos, e levar à interpretação equivocada de seus conceitos subjacentes. O trabalho segue com uma avaliação de modelos e modelagem de dados, propondo uma adaptação do MEDS como método para a aquisição de conceitos no processo de modelagem de dados.

Este trabalho segue, então, na pesquisa dos processos de cognição humanos e da linguagem, dois aspectos cognitivos intimamente ligados. Além disso, o trabalho inclui ainda um estudo de conceitos da Semiologia e da Lingüística com vistas a compreender os mecanismos de representação de conceitos inatos ao homem e aplicar esses conhecimentos na avaliação de uma linguagem de representação capaz de expressar, com riqueza semântica, os conceitos levantados na modelagem.

10. Referências

- Batini, C., Ceri, S. e Navathe, S. (1992) , *Conceptual Database Design*, The Benjamin/Cummings Publishing Company.
- Berners-Lee, T. (2008) “Sir Tim Berners-Lee Talks with Talis about the Semantic Web”
http://talis-podcasts.s3.amazonaws.com/twt20080207_TimBL.html
- Chandler, D. (1995) “Semiotics for Beginners”. Disponível em:
<http://www.aber.ac.uk/media/Documents/S4B/semiotic.html>
- Chen, P. (1990), *Modelagem de Dados – A Abordagem Entidade-Relacionamento para Projeto Lógico*, Makron Books, 1ª edição.
- Elmasri, R., Navathe, S. (2007), *Fundamentals of Database Systems*, Pearson Addison Wesley, 5ª edição.
- Guizzardi, G. (2005), *Ontological Foundations for Structural Conceptual Models*, CTIT.
- Heuser, C. (2001), *Projeto de Banco de Dados*, Sagra Luzzato, 1ª edição.
- Hitchman, S. e Veres, C. (2002) “Using Psychology to Understand Conceptual Modelling”. Disponível em: <http://is2.lse.ac.uk/asp/aspecis/20020014.pdf>
- Jackendoff, R. (1983), *Semantics and Cognition*, MIT, 1ª edição.
- Kent, W. (1998), *Data and Reality*, 1st Books Library, 2ª edição.
- Nicolaci-da-costa, A. (2007) “O Campo da Pesquisa Qualitativa e o Método de Explicitação do Discurso Subjacente (MEDS)”. Em: <http://www.scielo.br/prc>
- Saussure, F. (2006), *Curso de Lingüística Geral*, Cultrix, 28ª edição.
- Azevedo, L.; Siqueira, S.; Baiao, F.; Souza, J.; Lopes, M.; Cappelli, C.; Santoro, F. (2009) “Enterprise Ontology Management: An Approach Based on Information Architecture”. In: *Int Conf on Enterprise Information Systems (ICEIS)*, Springer
- Sternberg, R. (2008), *Psicologia Cognitiva*, Artmed, 4ª edição.
- Férrandez, M., Gómez-Pérez, A., Juristo, N. (1997) "METHONTOLOGY: From Ontological Art Towards Ontological Engineering". *AAAI97 Spring Symposium Series on Ontological Engineering*, pp. 33-40, EUA.
- Orlean, D. (2003) “Um Processo Unificado para Engenharia de Ontologias”, *Dissertação de Mestrado*, Departamento de Informática, PUC-Rio de Janeiro.
- Noy, N., McGuinness, D. (2001) “Ontology development 101: A guide to creating your first ontology”, *Relatório Técnico KSL-01-05*, Stanford