

## ONTOLOGIAS E BANCOS DE DADOS

Fabrizio Ricardo Lazilha; Flávio de Souza Bernabé; Márcia Cristina Debalto Pascutti

Faculdades Integradas de Maringá  
Centro de Ensino Superior de Maringá

**RESUMO:** Este artigo conceitualiza "Ontologia" como um conjunto de termos cuja finalidade é padronizar e axiomatizar a manutenção das informações de forma compartilhada e reutilizável por diferentes usuários. O principal propósito de uma ontologia é tornar explícita a informação de maneira independente das estruturas de dados subjacentes que podem ser usadas para armazenar a informação em depósitos de dados. Ontologias são abstrações e podem descrever diferentes tipos de organização de dados como tabelas relacionais, textos e imagens. Usuários deveriam ser capazes de interagir com a ontologia ao invés de interagir com vários depósitos de dados heterogêneos. Assim, usuários formulariam consultas sobre a ontologia e o sistema teria a responsabilidade de gerenciar a heterogeneidade e distribuição nos depósitos, isto é, uma ontologia define uma "linguagem" que será usada para formular consultas..

**Palavras-chave:** ontologia; banco de dados; axiomatização; compartilhamento; reutilização.

**ABSTRACT:** This paper conceptualize Ontologies as a group fo terms whose purpose is to standardize and axiomatize the manipulation of the information in a shared was for different users. The main purpose of an ontologie is to turn explicit the information in independent was of the structures of underlying data that can as tables you relate, texts and images. Users should be capable of interact with the ontologie instead of interact with several deposits of heterogeneous data. Thus, users would formulate consultations on the ontologie and system would have to manage the responsibility the heterogeneity and distribution in the deposits, that is, an ontologie defines a language that will be used to formulate consultation.

**Keywords:** ontologies.

### Introdução

O volume de informações e a variedade de estruturas de dados empregadas tem crescido exponencialmente nos últimos anos. A constante necessidade de consultas a esses dados (Sistema de Informações Global) tem estimulado o desenvolvimento de técnicas de consultas capazes de tornar explícita a informação de maneira independente das estruturas de dados usadas. Estas técnicas baseiam-se na construção de uma visão semântica única, uma ontologia.

Uma ontologia define uma "linhagem" que será usada para formular consultas. Ontologias são abstrações e podem descrever diferentes tipos de organização de dados como tabelas relacionais, textos e imagens.

Usuários deveriam ser capazes de interagir com a ontologia ao invés de interagir com vários depósitos de dados heterogêneos. Assim, usuários formulariam consultas sobre a ontologia e o sistema teria a responsabilidade de gerenciar a heterogeneidade e distribuição nos depósitos.

O desenvolvimento de ontologias é mais propriamente uma arte do que uma ciência. A causa disso é que não há nenhuma definição e padronização de um ciclo da vida, metodologias e técnicas que dirigem o desenvolvimento de ontologias.

Atualmente, há um grande interesse no desenvolvimento de ontologias para facilitar o compartilhamento de informações em geral e integração de bancos de dados heterogêneos e distribuídos.

Este artigo está estruturado da seguinte maneira: a seção dois trata da conceitualização de ontologias relacionadas a integração de bancos de dados. A seção 3

trata do uso ontologias como visões semânticas. A seção 4 descreve o OBSERVER, um gerenciador de consultas através de ontologias, proposto por NIETO (1998) em sua tese de doutorado. A última seção apresenta algumas aplicações de ontologias descritas por alguns autores com publicações relacionadas à ontologias.

### Conceito de Ontologia

O Dicionário Michaelis trata ontologia como sendo a *ciência do ser em geral*, ou ainda como a *parte da metafísica que estuda o ser em geral e suas propriedades transcendentais*.

Já a Filosofia, segundo REZENDE (1986) define ontologia como sendo parte da filosofia que estuda o *ser enquanto ser*, isto é, independentemente de suas determinações particulares e naquilo que constitui sua inteligibilidade própria; ou ainda como sendo a *teoria da intuição racional do absoluto como fonte única, ou pelo menos principal do conhecimento humano*.

Segundo GRUBER (1993),

*Ontologia é uma especificação de uma conceitualização, quer dizer, uma ontologia é uma descrição (como uma especificação formal de um programa) dos conceitos e relacionamentos que podem existir para um agente ou uma comunidade de agentes.*

Esta definição serve apenas para especificar um conjunto de conceitos. O importante em uma ontologia é sua aplicação. O autor tem projetado ontologias com a finalidade de permitir o compartilhamento de informações e seu reuso.

Neste contexto, uma ontologia é uma especificação usada para definir compromissos ontológicos.

Por razões pragmáticas, GRUBER (1993) escolheu escrever uma ontologia como um conjunto formal de definições de vocabulário. Embora este não seja o único modo para especificar uma conceitualização, ele tem algumas propriedades úteis para compartilhar informações entre softwares de inteligência artificial (por exemplo, semântica independente do leitor e contexto). Praticamente um compromisso ontológico é um acordo para usar um vocabulário (isto é, requer consultas e fazer afirmações) de maneira consistente com a ontologia específica. Agentes são construídos para garantir esses compromissos. Ontologias são projetadas para que informações possam ser compartilhadas com e entre esses agentes.

Uma das muitas utilidades da ontologia tem sido a modelagem de problemas e domínios em áreas como reengenharia de processos de negócios (onde se precisa de um modelo integrado da empresa e seus processos, sua organização, suas metas, e seus clientes), em arquiteturas multi-agentes distribuídas (onde diferentes agentes precisam comunicar-se e resolver problemas), e em engenharia concorrente a projeto.

GRÜNINGER (1996) usa ontologias para modelagem de problemas e domínios. Ontologias são planejadas para facilitar a reusabilidade de bibliotecas de objetos para modelagem de problemas e domínios. A última meta desta técnica é a construção de uma biblioteca de ontologias as quais podem ser informais, pois não provém semânticas nem axiomatizações. Nós precisamos de um modelo para especificar a semântica dos objetos e suas relações na ontologia, afirma Grüninger.

### Uso de ontologias

Numa arquitetura que permite processamento inteligente de consultas em um Sistema de Informações Global, a informação disponível nos diferentes depósitos deveria ser descrita por visões semânticas. De fato, cada depósito de dados deveria ser descrito por pelo menos uma visão semântica. Assim, a heterogeneidade e distribuição existente entre os depósitos de dados no Sistema de Informações Global estariam escondidas dos usuários que poderiam trabalhar com um número pequeno de visões semânticas.

Ontologias têm sido aceitas como poderosas ferramentas de descrição e, por esta razão, são muito apropriadas para representar o papel de visões semânticas (ver Fig. 1).

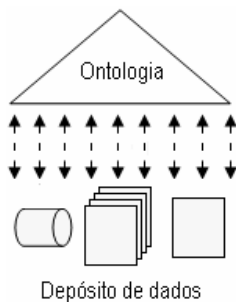


Figura 1: Ontologia descrevendo depósitos de dados

Atualmente, há um grande interesse no desenvolvimento de ontologias para facilitar o compartilhamento de informações em geral e integração de bancos de dados. Uma ontologia é um conjunto de condições de interesse em um domínio de informações particulares e suas relações. Ontologias e relacionamento entre ontologias são criados por especialistas no domínio correspondente, mas eles podem representar um ponto-de-vista particular do Sistema de Informações Global, isto é, podem descrever domínios personalizados.

Em seu artigo, NIETO (1998) salienta que ontologias são definidas usando descrições lógicas. Para permitir um processamento inteligente de consultas, o autor descreve qualquer depósito de informações por uma ou mais ontologias que está relacionada a dois depósitos de dados.

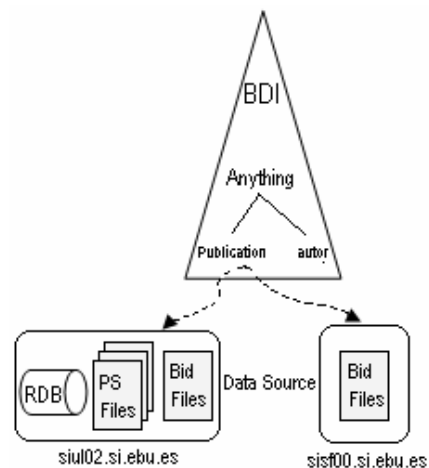


Figura 2: ontologias como ferramentas para descrever depósitos de dados.

Observando a ontologia (Fig. 2), percebemos qual o tipo de informação está armazenada no depósito de dados subjacente (informações relacionadas para publicações e seus autores). Depósitos são bases de dados relacionais e um conjunto de arquivos distribuídos entre dois computadores. Este é exatamente o principal propósito. Assim, usuários formulam consultas sobre ontologias e o sistema tem a responsabilidade de gerenciar a heterogeneidade e distribuição nos depósitos, isto é, uma ontologia define uma linguagem (conjunto de termos) que pode ser usada para formular consultas.

Além disso, ontologias podem ser vistas como compromissos entre provedores de informação (que definem ontologias e seus mapeamentos para depósitos de dados subjacentes) e usuários de informações. Consequentemente, usuários avançados ou organizações poderiam criar suas próprias ontologias, não para descrever seus próprios depósitos, mas para descrever seus pontos de vista sobre o Sistema de Informações Global; o sistema de informação é então visto como somente um depósito de dados.

Desde que ontologias constituem um modelo semântico sobre os depósitos de dados que eles descrevem, termos na ontologia podem representar qualquer informação, existente ou não no depósito em um momento específico (isto é, uma ontologia pode descrever que há livros em um

determinado depósito embora é possível que livros não tenham sido armazenados no depósito). Como ontologias são abstrações, elas também podem descrever qualquer tipo de formato de dados, desde textual até objetos multimídia.

Embora seja possível obter uma integração global de ontologias (descrevendo todo o Sistema de Informações global), NIETO (1998) defende o procedimento com múltiplas ontologias porque o gerenciamento de ontologias globais integradas envolve problemas de administração, manutenção, eficiência e consistência, muito difíceis de resolver. Uma ontologia muito grande também pode ser muito difícil para um usuário navegar e compreender. Também, não é realidade assumir que uma ontologia global simples possa descrever todos os dados disponíveis na Web, por exemplo. Além do mais, uma ontologia integrada global força todos os usuários a utilizar seu vocabulário, assim, diferentes ontologias descritas usando diferentes vocabulários podem satisfazer as necessidades do usuário de uma melhor maneira, reduzindo problemas de consistência e eficiência (ver Fig. 3).

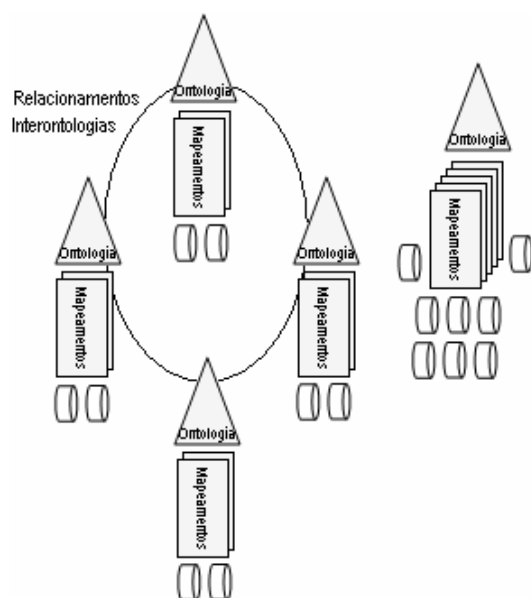


Figura 3: Múltiplas ontologias X

Porém, diferentes ontologias não são completamente ortogonais. Nem é provável que a necessidade de informação de um usuário seja satisfeita pelo acesso ao depósito de dados através do mapeamento associado à ontologia simples. Para suportar isto, ontologias são ligadas por relacionamentos interontologias. Esses relacionamentos podem ser usados para dois propósitos: o primeiro, para outra, e o segundo, para indiretamente suportar processamento de consultas que acessariam dados descritos por múltiplas ontologias.

Ontologias que compartilham uma grande similaridade podem ser organizadas em blocos para facilitar a escolha da ontologia mais apropriada. Blocos representam áreas gerais de conhecimento como *animais*, *bibliotecas*, *artes*, etc. De qualquer forma, podem existir, e isto é muito freqüente, relacionamentos entre ontologias em diferentes blocos, da mesma maneira que especialistas em áreas específicas algumas vezes precisam de informações

gerenciadas por pessoas em diferentes áreas. Desde que alguns blocos possam ser mais gerais que outros, eles poderiam ser organizados em hierarquias para facilitar a seleção do bloco mais apropriado.

Do ponto de vista de NIETO (1998), uma ontologia é um conjunto de termos de interesse num particular domínio da informação e nos relacionamentos entre eles. Ontologias e seus relacionamentos são criados por especialistas no domínio correspondente. Eles podem também representar um particular ponto de vista do Sistema de informações Global, isto é, eles podem prescrever domínios personalizados de usuários avançados.

Existem duas alternativas para a tarefa de fazer encapsulamento semântico de depósitos de dados por ontologias: 1) definir novas ontologias, ou 2) reusar ontologias pré-existentes. Ontologias são criadas para descrever depósitos de dados mas é também possível usar ontologias pré-existentes para encapsular novos depósitos de tais ontologias descrevem semanticamente aqueles novos depósitos. Somente quando uma ontologia não se ajuda a semântica de um depósito, uma nova ontologia deveria ser descrito por diversas ontologias as quais o "vêem" de diferentes pontos de vistas.

#### Processamento de Consultas Formuladas através de Ontologias

NIETO (1998), em sua tese de doutorado, define e implementa o OBSERVER, uma técnica para se processar uma consulta através de ontologias, deve-se seguir 3 passos: construção da consulta, acesso aos dados subjacentes e a expansão da consulta para outras ontologias. Três módulos são encarregados de executar estes três passos: o *Processador de consultas*, o *Gerenciador de Relacionamentos Interontologia* e o *Servidor de Ontologias*.

1. *construção de Consulta*. Um usuário em algum momento tem necessidade de algum tipo de informação. Assim, pela conexão com o *Processador de consulta* e usando a GUI este usuário pode construir uma consulta para a qual o sistema providenciará uma resposta. O processador de Consulta mostra ao usuário diversas áreas de conhecimento (que correspondem aos blocos de ontologia: esportes, notícias, ciência, biblioteca, etc.) para restringir o domínio semântico ao qual a consulta está relacionada. A informação sobre áreas de conhecimento e ontologias em cada área é obtida pelo Processador de Consulta depois de consultar a informação gerenciada pelo Gerenciador de Relacionamentos Interontologia (GRI) (este módulo representa o papel do catálogo do sistema e armazena informações semânticas entre ontologias). Quando o usuário escolhe uma área de conhecimento o Processador de Consulta apresenta uma visualização 3D contendo as ontologias no sistema. O usuário pode então navegar as ontologias e também consultar descrições de seus termos organizados para escolher a ontologia que melhor se ajusta às suas necessidades de informação.

Depois de escolher uma ontologia, o usuário constrói uma consulta que tem que ser executada pelo Processador de Consulta. A intervenção do usuário não é mais necessária. A ontologia escolhida será chamada aqui como *ontologia do usuário* para denotar que foi a ontologia usada para formular a consulta.

2. *Acesso aos Dados Subjacentes.* O Processador de Consulta recupera primeiro o dado correspondente à consulta que reside sob a ontologia do usuário. Para esta tarefa, o processador de consulta chama o *Servidor de Ontologia* correspondente para a ontologia do usuário (em cada ponto que contém ontologias há um Servidor de Ontologia que provê acesso a elas e seus depósitos de dados subjacentes onde quer que eles estejam).

O Servidor de Ontologia usa a informação mapeada correspondente para traduzir a consulta do usuário em diferentes consultas para os depósitos de dados subjacentes (cada sub-consulta é expressa numa linguagem de consulta do depósito onde será executada). O Servidor de Ontologia conhece quais depósitos de dados estão relacionados com a ontologia do usuário, em quais pontos eles residem, o tipo de organização de dados, e como acessa-los. Deste modo, o Servidor de Ontologia recupera o dado correspondente para a consulta. Dados são recuperados de diferentes depósitos num formato comum para que as sub-respostas diferentes sejam combinadas (correlacionadas). A resposta correspondente é retomada para o Processador de Consulta que a apresenta ao usuário. Se o usuário está satisfeito com a resposta, finaliza-se o processo de consulta.

3. *Expansão das Consulta para Outras Ontologias.* Se o usuário deseja mais dados relevantes então outras ontologias devem ser visitadas. Assim, a consulta original tem que ser traduzida dos termos da ontologia do usuário para os termos da outra ontologia componente (chamada ontologia alvo). Como a tradução pode não ser exata (nem todas as abstrações representadas na ontologia do usuário aparecem nas demais ontologias do sistema), o usuário pode definir um limite permitido para a perda de informações. Por exemplo, e o usuário define um limite de 20%, o sistema garante que a quantia de dados indesejados ou perdidos nas futuras respostas apresentadas para o usuário mantenham-se sempre abaixo de 20% das informações apresentadas. Naturalmente, o usuário pode definir um limite de 0% para garantir que a resposta sempre coincida exatamente com a semântica da sua consulta.

Para a tarefa de tradução da consulta nos termos da ontologia do usuário para os termos da ontologia alvo, o usuário e a ontologia alvo são integrados automaticamente pelo Processador de Consulta após consultar a semântica de relacionamentos interontologia que foi definida entre o usuário e a ontologia alvo. A informação que semanticamente relata diferentes ontologias é definida quando novas ontologias são acopladas ao Sistema de Informações Global ontologias são armazenadas e gerenciadas pelo GRI. No processo de integração, a consulta do usuário é reescrita (talvez parcialmente) nos termos da ontologia alvo. Se isto resulta em uma tradução completa, o Processador de consulta chama o Servidor correspondente à ontologia alvo o qual recuperará dado subjacente que corresponde à consulta traduzida.

Se após o processo de integração o Processador de Consulta obter uma tradução parcial, ele tenta combiná-la com outras traduções parciais obtidas anteriormente para alcançar uma nova tradução completa. Se isso não é possível, o sistema tenta substituir os termos em conflito (aqueles sem tradução na ontologia alvo) por outras expressões para obter uma tradução completa (isto é, uma nova consulta onde todos os termos são da ontologia alvo),

mas com uma perda da informação associada. Esta tarefa pode gerar muitas alternativas de tradução, para cada uma o Processador de Consulta estima perda de informação associada, rejeitando aquelas que incorrem em mais perda que o limite estabelecido pelo usuário. Aquela com menor perda é escolhida e enviada para o Servidor de Ontologia para obter os dados subjacentes correspondentes. A nova resposta é relacionada com as anteriores e o resultado (e as perdas associadas) é apresentado para o usuário que pode escolher novamente entre terminar o processo ou enriquecer a resposta através de visitas a novas ontologias. Neste último caso, a nova ontologia alvo é escolhida e o mesmo algoritmo é executado novamente até que o usuário esteja satisfeito, como mostrado na Figura 4.

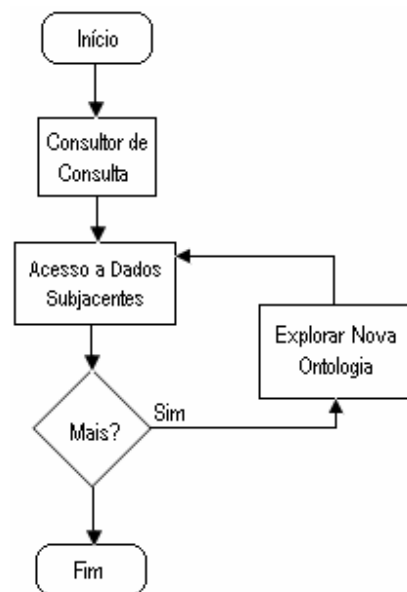


Figura 4: OBSERVER: Técnica de Processamento de Consulta.

Segundo os autores do artigo *Toward a Method to Conceptualize Domain Ontologies* (Rumo a um Método para Conceitualização de Ontologias) GOMEZ PEREZ (1996), o desenvolvimento de ontologias requer a definição e padronização de um ciclo de vida que parte das definições dos requisitos de um ciclo de vida que parte das definições dos requisitos até a manutenção do produto acabado, como também metodologias e técnicas que conduzem o seu desenvolvimento. Normalmente, cada grupo de desenvolvimento segue um conjunto de princípios, critérios, e fases no processo de desenvolvimento de ontologia. A ausência de diretrizes e métodos torna difícil o desenvolvimento de ontologias padronizadas, dentro e entre grupos, o complemento de uma determinada ontologia por outros grupos, e seu reuso em outras ontologias e aplicações finais. É proposto o uso em outras ontologias e aplicações finais. É proposto o uso de um modelo conceitual comum e compartilhado como ponto de partida, distinguindo quanto passos comuns no desenvolvimento de qualquer ontologia: a captura de conhecimento, sua conceitualização em um conjunto de técnicas, sua implementação em uma linguagem formal, e sua avaliação durante cada fase e entre as fases de seu ciclo de vida. O método usa um conjunto de técnicas na

fase conceitual que permite identificar conceitos, seus atributos e valores, bem como as relações entre eles. Para provar a viabilidade das idéias propostas no artigo, os autores conceitualizam a ontologia no domínio de substâncias químicas.

A metodologia proposta no artigo *Designing na Evaluating Generic Ontologies* (Desenvolvimento e Avaliação de Ontologias Genéricas) GRÜNINGER (1996) apóia a formalização de ontologias informais e a integração e reuso destas ontologias. A meta é escrever axiomas suficientes para capturar adequadamente a semântica pretendida. O desafio fundamental é dar alguma direção sobre os tipos de axiomas que um engenheiro ontológico precisa escrever.

O rtigo *Designing and Exploiting Na Ontology for Multi-Lingual Generation* (Projetando e Explorando uma Ontologia para Geração Multi-Lingual) HORACEK (1997) contribui com uma técnica sistemática para construir ontologias e traçar componentes numa representação léxica a qual pode estruturalmente definir através da linguagem para especializações ontológicas idênticas.

Em sua tese de doutorado, NIETO (1998) define e implementa o sistema OBSERVER, que permite aos usuários formular consultas ao Sistema de Informações global, cuidando somente da semântica dos dados nos quais estão interessados. Aspectos relacionados a sintaxe, localização, estrutura, organização dos dados, etc., são manipulados pelo sistema. O sistema é ou está baseado na idéia de usar ontologias como metadados para descrever depósitos de dados heterogêneos e distribuídos, assim a primeira principal contribuição consiste em prover um mecanismo que permita isto. Além disso, ontologias são consultadas por usuários e, então, a segunda contribuição consiste em prover um mecanismo que permite obter as respostas acessando depósitos de dados subjacentes.

### Conclusão

Atualmente, há um grande interesse no desenvolvimento de ontologias para facilitar o compartilhamento de informações em geral e3 integração de bancos de dados heterogêneos e distribuídos (Web). Neste artigo conceitualizamos "Ontologia" como um conjunto de termos cuja finalidade é padronizar e axiomatizar a manipulação das in formações de forma compartilhada e reutilizável por diferentes usuários.

### Referências

- REZENDE, A. *Curso de Filosofia*. Rio de Janeiro: Jorge Zahas, SEAF, 1986.
- GÓMEZ-PEREZ, Asunción et al. *Towards a Method To Conceptualize Domain Ontologies*. ECAI '96 Workshop Ontological Engineering, 1996. <http://www.cs.utwente.nl:8080/kbs/ecaiworkshop/fullpapers.html>
- GRÜNINGER, Michael. *Designing and Evaluating Generic Ontologies*. ECAI '96 Workshop Ontological Engineering, 1996. <http://wwwis.cs.utwente.nl:8080/kbs/ecaiworkshop/fullpapers.html>
- HORACEK, Helmut. *Designing and Exploiting An Ontology for Multi-Lingual Generation*. IJCAI '97 Workshop

WP243 Ontologias and Multi-Lingual NLP  
<http://crl.nmsu.edu/events/ijcai>

NIETO, Eduardo Mena. *OBSERVER: An Approach For Query processing In Global. Information Systems based on Interoperation across Pre-existing Ontologies*. Tesis Doctoral Universidad de Zaragoza. Novembro, 1998.

GRUBER, T.B. *Translation Approach to Portable Ontologies*, 1993. <http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>

