Uma Ontologia para o Controle Cambial Brasileiro: estudo de caso.

Aluno : Robson Moron
Orientador: Flavio Soares Correa da Silva

Sumário

Este trabalho apresenta o resultado dos estudos realizados no segundo semestre de 2002 na disciplina MAC5701 - Tópicos em Ciência da Computação, com vistas à elaboração da dissertação de mestrado " Uma Ontologia para o Controle Cambial Brasileiro: estudo de caso". Começaremos com a definição de Ontologia: uma especificação de uma conceituação, detalhando o que é conceituação e especificação para esta definição. A seguir mostraremos alguns ambientes de desenvolvimento de ontologias, e finalmente destacaremos algumas metodologias para ontologias implementadas.

Introdução

Após o surgimento do processamento de dados através dos computadores das empresas, os processos que eram executados manualmente passaram a ser automatizados e apenas esse fato aumentou consideravelmente a produtividade e a lucratividade das empresas que aderiram a tecnologia de informação(T I). Este aumento de produtividade e lucratividade fez com que ocorresse uma rápida e continua automatização dos processos. Infelizmente isso ocorreu sem o devido planejamento, gerando uma dualidade em T I, pois embora os Centros de Processamento de Dados aumentem a produtividade da empresa, eles possuem alguns problemas que fazem com que sua própria produtividade fique abaixo do desejável.

Um desses problemas é o custo gerado para o desenvolvimento e manutenção de sistemas. Mesmo desenvolvendo vários sistemas para um mesmo domínio, atualmente, não existe um aproveitamento de toda a estrutura já definida em sistemas anteriores.

A Ontologia permite que seja criada uma estrutura comum para um determinado domínio, fazendo com que exista a possibilidade de reutilização para cada sistema gerado no mesmo domínio, resultando assim em um ganho de produtividade e confiabilidade para o CPD.

A Ontologia também facilita o compartilhamento de informação, pois possibilita a geração de especificações que podem ser interpretadas por máquinas e posteriormente utilizadas por outras aplicações. A Web Semântica é um bom exemplo de utilização.

O que é Ontologia

No dicionário encontramos que Ontologia[1] é o tratado dos seres em geral, parte da metafísica que estuda o ser em geral e suas propriedades transcendentais. Entretanto O conceito mais conhecido de Ontologia em T I é que Ontologia[2] é uma especificação de uma conceituação. Existem outros conceitos interessantes sobre Ontologia, mas para o objetivo deste trabalho iremos admitir que Ontologia é uma especificação de uma conceituação.

Antes de descrevermos o que é uma conceituação e uma especificação para esta definição, vamos apresentar os tipos de Ontologias existentes, Guarino [10,11]:

Ontologias genéricas: descrevem conceitos bastante gerais, tais como, espaço, tempo, matéria, objeto, evento, ação, etc., que são independentes de um problema ou domínio particular;

Ontologias de domínio: descrevem conceitos de domínios particulares, descrevendo o vocabulário relacionado a um domínio, tal como Medicina e Direito.

Ontologias de tarefas: expressam conceituações sobre a resolução de problemas, independentemente do domínio em que ocorram, isto é, descrevem o vocabulário relacionado a uma atividade ou tarefa genérica, tal como vendas;

Ontologias de aplicação: descrevem conceitos dependentes do domínio e da tarefa particulares. Estes conceitos freqüentemente correspondem a papéis desempenhados por entidades do domínio quando da realização de uma certa atividade;

Ontologias de representação: explicam as conceituações que fundamentam os formalismos de representação de conhecimento.

Neste trabalho estaremos interessados em Ontologias de Domínio, pois o objetivo final é a elaboração da dissertação " Uma Ontologia para o Controle Cambial Brasileiro", assim, vamos detalhar os termos conceituação e especificação visando construir esta Ontologia de domínio.

Conceituação

Conceituação se refere a um modelo abstrato de algum fenômeno do mundo. Esta é uma definição muito ampla para a palavra conceituação, direcionando para uma Ontologia de domínio, a conceituação será um modelo abstrato de um domínio especifico. Para obter esta conceituação, necessitamos elaborar uma análise de domínio. O termo Análise de Domínio foi introduzido por Neighbors [9] com a seguinte definição:

A Análise de Domínio é uma tentativa de identificar os objetos, operações e relações entre o que peritos em um determinado domínio percebem como importante.

Esta atividade pode ser considerada praticamente equivalente à análise de requisitos tradicional, com a diferença que os requisitos explorados dizem respeito a uma família de aplicações de um determinado domínio.

Se considerarmos, por exemplo, o domínio do Controle Cambial Brasileiro:

Objetos típicos são Contratos de Câmbio, Declarações de Importação, Registros de exportação, Importadores, exportadores, moedas e mercadorias. Operações e ações incluem vínculos entre contratos de cambio e declaração de importação, contratos de câmbio e registro de exportação, CNPJ e declaração de importação. Uma linguagem específica de domínio pode ser criada para representar estes objetos, operações e relações, que pode, posteriormente, ser utilizada para descrever outros sistemas neste domínio.

Ressaltamos novamente que esta atividade pode ser considerada equivalente à análise de requisitos convencional, no entanto atuando em um meta-nível e, portanto, ao invés de explorar requisitos de uma aplicação específica, os requisitos explorados dizem respeito a uma família de aplicações de uma determinada área.

Segundo Asuncion Gómez Perez [5], Conceituação é importante em qualquer Engenharia de Software. A fase de conceituação abrange os objetos e suas relações, objetivando estruturar o domínio de conhecimento em um modelo conceitual. Uma metodologia ideal proporia duas atividades simultâneas na fase de conceituação, a análise e a síntese. A análise detecta o conhecimento concreto, discernível e estratégico e o meta conhecimento do domínio. A síntese utiliza esse conhecimento para construir modelos que levam ao modelo conceitual do sistema. Ambas atividades utilizam um conjunto de representações intermediarias (IR) para representar o conhecimento.

Outra definição esclarecedora e mais formal de Conceituação é dada por Nicola Guarino [11].

Seja a estrutura <D,W> tal que D representa o domínio em questão e W representa todos as possíveis estruturações de significado entre os elementos existentes em D. Esta estrutura é chamada de Espaço de Domínios. Como exemplo, D pode ser o conjunto de blocos sobre a mesa e W é um conjunto máximo de estado de ocupação do domínio(também chamado de mundos possíveis).

Uma conceituação C é uma estrutura <D,W, Â> sendo Â o conjunto de relações escolhidas como pertinentes para representar este domínio. Desta forma, uma conceituação define uma estrutura pretendida do mundo, representada por S.

Através das definições de Conceituação apresentadas, verificamos que várias conceituações podem ser feitas para um mesmo domínio, portanto a escolha da semântica dada a conceitos e relações será determinada pela natureza do universo de problemas que se tem em mente e pelas pessoas que irão formaliza-las, tornando a conceituação também um processo político, já que dificilmente uma conceituação será totalmente adequada a todos.

Especificação

Segundo Mike Uschold [13], uma especificação pode ter vários graus de formalidade, conforme resumo a seguir:

- Informal - "linguagem natural", gerando Ontologias que são utilizadas e compreendidas apenas por pessoas.

- Semi-informal - "expressa em uma forma estruturada e restrita de linguagem natural". Exemplo: a versão de "Enterprise Ontology"em formato texto.

- Semi-formal - "adota uma linguagem formalmente definida com sintaxe-rígida", temos como exemplo versão de "Enterprise Ontology"em Ontolingua.

- Rigorosamente formal - "teoremas e provas".

Neste trabalho aprofundaremos nossos estudos em Ontologias com especificação semi-formal, para isso apresentaremos agora XML, RDF e qual a importância disso para a WEB Semântica.

XML (*eXtensible Markup Language*)

A linguagem XML (*eXtensible Markup Language*) foi desenvolvida pelo W3C (*World Wide Web Consortium*) como um perfil de SGML (*Standard General Markup Language*) designado principalmente para aplicações Web. A partir desta linguagem, é possível definir e criar extensões para linguagens de marcação. XML é uma metalinguagem que é utilizada para definir outras linguagens, tornando-se assim, a base para uma nova maneira de se comunicar através da Internet, alem de ser utilizada para criar outras linguagens de marcação como a Linguagem de Marcação da Plataforma Lattes. As marcas em XML assemelham-se muitas vezes as marcas em HTML, mas diferem desta pela capacidade de oferecer uma boa interoperabilidade.

Existem 10 regras básicas para o XML. Vamos ver quais são os objetivos definidos para o XML pelo W3C:

1 - Será diretamente utilizado na Internet.

2 - Será suportado por uma vasta variedade de aplicações.

3 - Será compatível com o SGML.

4 - Será fácil escrever programas que façam o processamento de documentos XML.

5 - O número de características opcionais no XML deverá ser mantida em um número mínimo(o ideal é zero).

6 - Os documentos XML deverão ser passiveis de ser lidos por humanos e razoavelmente claros.

7 - O desenho do XML deverá ser preparado rapidamente.

8 - O desenho do XML deverá ser formal e conciso.

9 - Os documentos XML deverão ser facilmente criados.

10 - A sintetização das marcas XML é de mínima importância.

Exemplo de utilização do XML

Mostraremos agora um exemplo da Linguagem de Marcação da Plataforma Lattes(LMPL)[15]. A LMPL foi implementada utilizando o XML(*eXtensible Markup Language*).

<?xml version="1.0"encoding="ISO-8859-1"?>

- <CURRICULO-VITAE SISTEMA-ORIGEM-XML="LATTES-OFFLINE»

+ <DADOS-GERAIS NOME-COMPLETO="Fulano de Tal"

NOME-EM-CITACOES-BIBLIOGRAFICAS="TAL, F. "

NACIONALIDADE="B "

CPF="98765432109"PAIS-DE-NASCIMENTO="Brasil"

UF-NASCIMENTO="DF"

CIDADE-NASCIMENTO="Brasília"

DATA-NASCIMENTO="19500716"

SEXO="MASCULINO"

N U M E R O -IDENTIDADE="1234567890"

ORGAO-EMISSOR="SSP"

UF-ORGAO-EMISSOR="DF"

DATA-DE-EMISSAO="20001231"

NOME-DO-PAI="José Fulano de Tal"

NOME-DA-MAE="Sicrana dos Anzóis de Tal"

PERMISSAO-DE-DIVULGACAO="SIM»

+ <PRODUCAO-BIBLIOGRAFICA>

+ <PRODUCAO-TECNICA>

+ <OUTRA-PRODUCAO>

+ <DADOS-COMPLEMENTARES>

</CURRICULO-VITAE>

RDF (*Resource Description Framework*)

Enquanto a linguagem XML define uma estrutura, RDF permite expressar o significado associado aos dados. O significado é expresso por conjuntos de triplos que podem ser escritos utilizando marcas XML. O primeiro elemento do triplo é o sujeito, o segundo o predicado e o terceiro é o objeto. O sujeito, o objeto e o predicado são identificados por URI (*Universal Resource Identifier*). A identificação do predicado permite a definição de um novo conceito ou verbo simplesmente pela definição de um apontador URI para o conceito. É muito importante utilizar URI's diferentes para conceitos diferentes ou para variações do mesmo conceito, como forma de garantir a consistência das inferências realizadas. Por exemplo, o conceito de endereço utilizado na linguagem corrente pode corresponder, na realidade, vários conceitos: pode ser um endereço de email ou de um endereço físico. É, pois, necessário que sejam definidos de formas distintas i.e. que sejam identificados por diferentes URI's.

A unicidade da relação conceito-URI garante a unicidade dos triplos de significado.

Desta forma, cada conceito tem uma identificação única e os triplos constituirão uma rede de informação relacionada na qual podem usar-se regras de inferência para fazer deduções a partir do significado dos dados definido pelo triplo.

Exemplo de utilização do RDF

Mostraremos agora um exemplo de documento RDF, sendo este documento uma parte do arquivo RDF gerado pelo Protégé-2000 para a Ontologia de Vinhos[14].

```
<?xml version=’1.0’ encoding=’ISO-8859-1’?>
<!DOCTYPE rdf:RDF[
<!ENTITY rdf ’http://www.w3.org/1999/02/22-rdf-syntax-ns#’>
<!ENTITY a ’http://protege.stanford.edu/system#’>
<!ENTITY wines ’http://protege.stanford.edu/wines#’>
<!ENTITY rdfs ’http://www.w3.org/TR/1999/PR-rdf-schema-19990303#’>]>
<rdf:RDF xmlns:rdf="&rdf;"
xmlns:a="&a;"
xmlns:wines="&wines;"
xmlns:rdfs="&rdfs;»
<rdf:Description rdf:about="&wines;Apple"
wines:name="Apple"
rdfs:label="Apple»
<rdf:type rdf:resource="&wines;Non-sweet fruit"/>
</rdf:Description>
```

Web Semântica

Algumas colocações sobre informação:

"O valor da informação aumenta com o uso."

"O valor da informação diminui com o tempo. A informação possui um tempo de vida útil, o qual é variável dependendo do tipo de informação."

"O valor da informação pode aumentar quando combinada com outra informação."

"A informação pode gerar outra informação."

Em relação a Web semântica e RDF:

"Web Semântica é uma extensão da Web atual em que a informação é dada com significado bem definido habilitando melhor computadores e pessoas para trabalhar em cooperação. "Tim Berners Lee, James Hendler e Ora Lassila.

"RDF é a base para o processamento de metadados, que providencia a interoperabilidade entre aplicações, que trocam informações compreensíveis na Web."Ora Lassila.

Verificamos que ambientes para construção de Ontologias, como o Protégé-2000 , geram a especificação em vários formatos, entre eles RDF e RDFS, portanto a mesma forma utilizada para a comunicação na Web semântica.

O que muda?

O leitor pode estar propenso a concluir que Ontologia é a especificação de uma análise de domínio. Mostraremos agora uma parte da comparação entre Ontologia e Análise de Domínio realizada por Giancarlo Guizzardi [8] em sua dissertação de mestrado na Universidade Federal de Vitória:

"Na construção de Ontologias, além do modelo conceitual e do vocabulário de termos, o modelo de domínio é composto de um conjunto de axiomas. Esses axiomas representam de maneira formal o conhecimento do domínio, permitindo a derivação de novos conhecimentos a partir dos já formalizados, a verificação da consistência e validação automática desses conhecimentos, a restrição da interpretação da semântica de conceitos e relações e, principalmente, a possibilidade de geração automática de infra-estruturas computacionais reutilizáveis a partir do modelo de domínio. Uma vez especificados, os axiomas formais devem ser formalizados em uma linguagem capaz de servir aos propósitos da Ontologia e/ou da atividade de engenharia de domínio e, ao mesmo tempo, incorporar a menor quantidade possível de compromissos ontológicos adicionais.

Apesar de todas as vantagens citadas, a principal contribuição dos axiomas formais da Ontologia ainda é a possibilidade de representação explícita do conhecimento do domínio em um nível de significação (nível ontológico) ao contrário do nível de estruturação (epistemológico), que é o máximo alcançado quando se tem apenas os outros dois modelos citados. A figura 1 é usada, a seguir, para exemplificar esta diferença. Nela, é apresentado um modelo de entidades e relacionamentos, mostrando uma pequena parte de um sistema acadêmico para uma universidade.

Esta representação de nível epistemológico mostra os conceitos aluno, disciplina e turma, e os relacionamentos de matrícula, histórico e oferta entre instâncias desses conceitos. Pode-se perceber que, neste nível, só é possível representar a estrutura das coisas no mundo. Entretanto, algumas questões importantes permanecem em aberto, entre elas: Que elementos do conjunto de entidades aluno podem estar relacionados com uma instância específica de turma? O que é exatamente uma instância de turma? Que subconjunto do produto cartesiano das instâncias de aluno e disciplina caracteriza precisamente o relacionamento histórico? Tais questões não são respondidas por esse modelo, já que ele é um modelo do nível epistemológico. A solução normalmente adotada nestes casos consiste em embutir este conhecimento dentro do código de um programa, o que dificulta o seu reuso e compartilhamento.

Construindo-se Ontologias formais, é possível estabelecer significados formais para alguns termos do vocabulário do domínio, assim como pode-se restringir a interpretação da teoria com base na axiomatização da Ontologia. No caso do sistema acadêmico, axiomas da Ontologia poderiam, por exemplo, impor condicionantes para

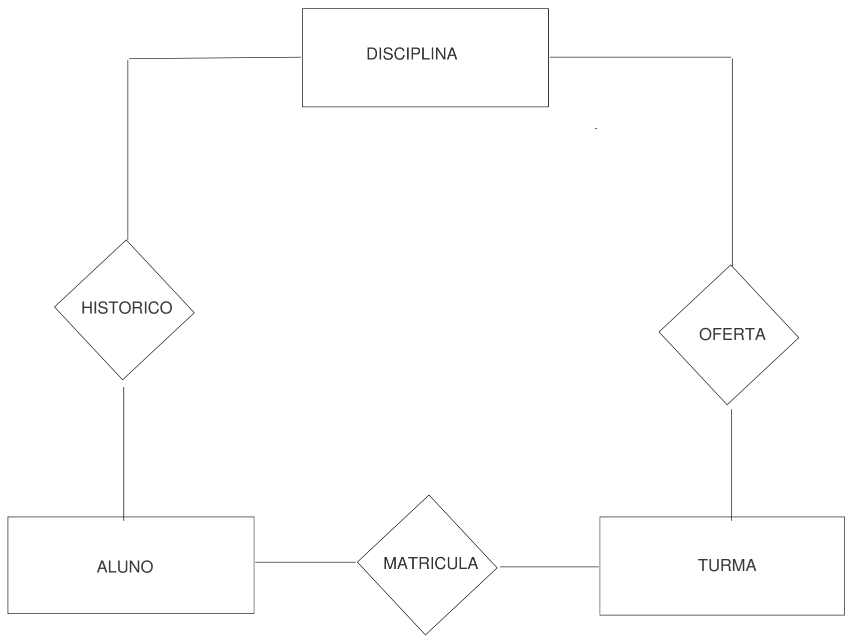

Figura 1: Extrato de um modelo ER para um Sistema Acadêmico.

o estabelecimento do relacionamento matrícula, tal como: se a é um aluno e existe um relacionamento histórico(a,d), indicando que este aluno já cursou a disciplina d tendo sido aprovado, então não é possível haver o relacionamento matrícula(a,t), se a turma t for uma oferta da disciplina d, (isto é oferta(t,d)) já cursada pelo aluno. A seguinte sentença, em lógica de primeira ordem, é a formalização deste axioma:

$(\forall\ a,d,t)\ (aluno(a\ ) \wedge disciplina(d) \wedge turma(t) \wedge$

$hist\acute{o}rico(a,d,Aprovado) \wedge oferta(t,d) \rightarrow \neg\ matr\acute{i}cula(a,t))$ "

Os ambientes para a construção de Ontologias, os quais serão apresentados a seguir, possuem interfaces amigáveis para especificação destes axiomas.

Ambientes para construção de Ontologias

Vários grupos de pesquisa estão empenhados em desenvolver ambientes visando a construção de Ontologias, dentre estes ambientes temos:

Apollo: http://apollo.open.ac.uk

Ontolingua: http://ontolingua.stanford.edu

OntoSaurus : http://www.isi.edu/isd/ontosaurus.html

Protégé-2000 : http://protege.stanford.edu

WebODE : http://webode.dia.fi.upm.es

Obs: Apolo e Protégé-2000 estão disponíveis para *Download*. Ontolingua e OntoSaurus (interface Gráfica do LOOM) podem ser utilizados via internet. WebODE pode ser utilizado via internet, mas o usuário deverá possuir um IP fixo.

Utilizamos o Protégé-2000 para o exemplo da Ontologia de vinhos e comidas [14] (Ontologia Introdutória), que especifica a conceituação do domínio Vinhos e comida em RDF/RDFS e também em formato Clips. Para o desenvolvimento desta Ontologia, o grupo de *Stanford* sugeriu sete passos, sendo os passos de um à três para a fase de conceituação e os quatro passos restantes com a fase de especificação.

Para a comparação dos ambientes para construção de Ontologias sugerimos a leitura de "A survey on Ontology tools"[7], pois entendemos que a melhor escolha da ferramenta a ser utilizada esta diretamente ligada as propriedades do Domínio de sua Ontologia.

Metodologia

Embora alguns grupos estejam trabalhando para a elaboração de uma metodologia para construção de Ontologias, como "Methontology de Asuncion Gomez Perez"para a Ontologia "Chemicals"e "Building Ontologies: Towards a Unified Methodology de Mike Uschold"para a Ontologia "Enterprise", concordamos com Asuncion Gómez Perez [5] que diz que a Engenharia de Ontologia ainda esta a mais para Arte do que para a Ciência.

Dentre as Ontologias Brasileiras, destacamos a Linguagem de Marcação da Plataforma Lattes(LMPL), que é uma padronização consensual entre algumas universidades brasileiras sobre a representação dos currículos e produções cientificas dos professores. A LMPL foi implementada utilizando o XML(*eXtensible Markup Language*).

Conclusão.

Neste trabalho apresentamos que a Ontologia é uma ferramenta eficaz para compartilhar e reutilizar o conhecimento. Em relação a reutilização enfatizamos que esta é uma boa ferramenta para o aumento da qualidade e produtividade nos CPD. Em relação ao compartilhamento de conhecimento, descrevemos o valor da informação e exemplificamos uma real utilização(web semântica).

Durante os estudos, observamos que uma Ontologia pode utilizar algumas das seguintes tecnologias: Análise de Domínio, Sistemas Orientados a Objetos, Frames, Ferramentas da Qualidade(Padronização, Brainstorming, Benchmark, . . .), Linguagens de Marcação(XML e RDF), Base de Dados, Metadados, Taxonomia, . . . , e outras que vierem a existir. Por este motivo podemos reduzir erroneamente a Ontologia a uma destas Tecnologias, sendo necessário ressaltar a adaptabilidade que a Ontologia possui com estas tecnologias e com outras que venham a ser desenvolvidas(Ex: uma outra meta-linguagem que substitua o XML, ou ainda uma outra forma de especificação).

Por outro lado, devemos ficar atentos para o fato da Ontologia ter suas limitações em relação à reutilização e ao compartilhamento, pois a Ontologia necessita das tecnologias disponíveis atualmente, assim, recebe também como herança os conhecidos problemas destas tecnologias (padronização - inflexibilidade, Análise de Domínio - processo político, XML/RDF - interesses privados, . . .).

Observamos ainda que apesar dos esforços dos grupos de pesquisadores de Ontologia e a grande quantidade de trabalhos publicados na área, inexiste uma metodologia consensual para construção e avaliação de Ontologias.

Referências Bibliográficas

[1] F. Fernandes, C. P. Luft e F. M. Guimaraes :*"Dicionário Brasileiro Globo."*

[2] Gruber, T. R.:*"Toward Principles for Design of Ontologies Used For Knowledge Sharing."*

[3] Frank Van Harmelen:*"Ontologies´KISSES in standardization."*

[4] Asuncion Gomez Perez:*"Methontology."*

[5] Asuncion Gomez Perez:*"Towards a Method to Conceptualize Domain Ontologies."*

[6] Asuncion Gomez Perez:*"Building Ontologies at the Knowledge level using the Ontology Design Environment."*

[7] Asuncion Gomez Perez:*"A survey on ontology tools."*

[8] Giancarlo Guizzardi:*"Dissertação de mestrado,UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. 2000."*

[9] NEIGHBORS J.:*"Software Construction Using Components. 1981. Tese (Doutorado) - Universidade da Califórnia, Irvine, EUA."*

[10] GUARINO, N.:*"Understanding, building and using ontologies. Int. Journal Human-Computer Studies, v. 45, n. 2/3, fev./mar. 1997."*

[11] GUARINO, N.:*“Formal Ontologies and Information Systems. In: FIRST INTERNATIONAL CONFERENCE (FOIS), 1., 1998, Trento, Itália. Anais... Trento: IOS Press, 1998.”*

[12] Mike Uschold:*“Building Ontologies: Towards a Unified Methodology.”*

[13] Mike Uschold:*“Ontologies: Principles, Methods and Applications.”*

[14] Natalya F. Noy, Debora L. Mcguinness:*“Ontology Development 101: A guide to creating your first Ontology.”*

[15] Roberto Carlos dos Santos Pacheco e Vinicius Medina Kern:*“Uma Ontologia comum para a integração de bases de informação e conhecimento sobre ciência e tecnologia.”*

[16] *“Pesquisas na internet.”*