

RELATÓRIO TÉCNICO

**UM AMBIENTE QUE INTEGRA
SISTEMAS ESPECIALISTAS E
BANCO DE DADOS
BASEADOS NO MODELO E-R**

**Esther de Castro Pacitti
Pedro Manoel Silveira**

**NCE-24/90
Outubro/90**

**Universidade Federal do Rio de Janeiro
Núcleo de Computação Eletrônica
Caixa Postal 2324
20001 - Rio de Janeiro - RJ
BRASIL**

UM AMBIENTE QUE INTEGRA SISTEMAS ESPECIALISTAS E BANCO DE DADOS BASEADOS NO MODELO E-R

RESUMO

Sistemas Especialistas e Bancos de Dados encerram tecnologias com vários pontos em comum. O principal deles é que, nos dois casos, armazena-se e manipula-se conhecimento, seja na forma de fatos ou de regras de dedução. Ainda assim, a integração dessas duas tecnologias ainda é obscura e pouco explorada. Este artigo apresenta as bases para uma integração de SE e BD baseada na utilização de regras de produção com termos e literais mais abrangentes, que suportam elementos de bancos de dados. Utiliza-se como referência o Modelo Entidade-Relacionamento, devido à sua intuitividade e também por ser esta uma tecnologia largamente utilizada para modelagem de dados. O ambiente de integração aqui exposto está sendo correntemente implementado para uma *shell* baseada no paradigma Objeto-Atributo-Valor (O-A-V)

AN ENVIROMENT FOR EXPERT SYSTEMS AND DATABASE INTEGRATION BASED ON THE E-R MODEL

ABSTRACT

Expert Systems and Databases are technologies with quite a few points in common. Their main intersection is that both store and manipulate knowledge, either as facts or rules. Even so, the integration of these areas is in some sense unexplored and cases where this happens are not frequent. This article presents the basis for such integration, where rules with extended terms have been introduced in order to support database elements. The Entity-Relationship Model has been used as a database reference model, due to its widespread use and easy understanding. The environment here presented is currently being implemented with an expert system shell based upon the O-A-V paradigm for knowledge representation.

Um Ambiente que Integra Sistemas Especialistas e Bancos de Dados Baseados no Modelo E-R

ESTHER DE CASTRO PACITTI

PEDRO MANOEL SILVEIRA

Núcleo de Computação Eletrônica — UFRJ

Caixa Postal 2324, Cep 20001, Rio de Janeiro, RJ, Brasil

E-mail: ncd10123@ufrj.bitnet, pedro@ufrj.bitnet

RESUMO

Sistemas Especialistas e Bancos de Dados encerram tecnologias com vários pontos em comum. O principal deles é que, nos dois casos, armazena-se e manipula-se conhecimento, seja na forma de fatos ou de regras de dedução. Ainda assim, a integração dessas duas tecnologias ainda é obscura e pouco explorada. Este artigo apresenta as bases para uma integração de SE e BD baseada na utilização de regras de produção com termos e literais mais abrangentes, que suportam elementos de bancos de dados. Utiliza-se como referência o Modelo Entidade-Relacionamento, devido à sua intuitividade e também por ser esta uma tecnologia largamente utilizada para modelagem de dados. O ambiente de integração aqui exposto está sendo correntemente implementado para uma *shell* baseada no paradigma Objeto-Atributo-Valor (O-A-V) .

ABSTRACT

Expert Systems and Databases are technologies with quite a few points in common. Their main intersection is that both store and manipulate knowledge, either as facts or rules. Even so, the integration of these areas is in some sense unexplored and cases where this happens are not frequent. This article presents the basis for such integration, where rules with extended terms have been introduced in order to support database elements. The Entity-Relationship Model has been used as a database reference model, due to its widespread use and easy understanding. The environment here presented is currently being implemented with an expert system shell based upon the O-A-V paradigm for knowledge representation.

Este trabalho foi parcialmente financiado com recursos do projeto ESTRADA da SID Informática S.A..

1. INTRODUÇÃO

Tradicionalmente, sistemas que manipulam conhecimento em Inteligência Artificial (IA) provêem meios flexíveis para criar, examinar e modificar, através de uma máquina de inferência, coleções de estruturas simbólicas formando o que é denominado *Base de Conhecimento*. Pode-se observar que os princípios envolvidos nos componentes de um sistema baseado em conhecimento dizem respeito à base de conhecimento e ao mecanismo de inferência. Neste aspecto, há uma certa semelhança com as funções dos sistemas de banco de dados, e pode-se notar um grande interesse em se unir tecnologias e resultados das áreas de Inteligência Artificial e Banco de Dados (BD).

A área de pesquisa em Banco de Dados, já mais antiga, apresenta um sucesso prático visível. Os pontos centrais desta área estão relacionados primordialmente com os aspectos de implementação e às características operacionais dos sistemas de Banco de Dados. Nota-se que os modelos de dados utilizados em BD compartilham os objetivos almejados pelos esquemas de representação do conhecimento utilizados em Inteligência Artificial, pois na realidade ambos representam domínios de aplicação.

É inegável que a integração destas duas tecnologias traz benefícios para ambas. A contribuição fundamental da tecnologia de IA consistirá na obtenção de modelos semânticos mais ricos, isto é, um entendimento mais claro a respeito do conhecimento representado. Por outro lado, a contribuição da tecnologia de BD propicia a habilidade de se realizar acessos eficientes a bases de conhecimento robustas e melhor estruturadas, com auxílio de controles de concorrência, segurança e proteção de dados, acesso otimizado e gerenciamento de memória secundária.

Sistemas baseados em conhecimento e SGBD (Sistemas Gerenciadores de Banco de Dados) possuem características arquiteturais semelhantes. Ambos baseiam-se na escolha de estruturas apropriadas para representar o conhecimento e mecanismos eficientes para responder a perguntas ou deduzir novos conhecimentos a partir dos já existentes. No entanto, sistemas baseados em conhecimento enfatizam dedução e representações complexas, enquanto que em SGBD a ênfase se dá na eficiência de acessos e na manipulação de estruturas, que são persistentes e armazenam informações de um universo específico.

Neste contexto, observou-se que triplas O-A-V, utilizadas na representação do conhecimento na forma (o, α, c) , respectivamente, poderiam ser vistas como funções $\alpha(o) = c$. Dentro desta ótica, intui-se que estas funções permitiriam acesso a banco de dados, desde que o objeto o fosse considerado um elemento de um conjunto de entidades ou relacionamentos, α um atributo, e c o valor do atributo em questão. Dentro desta filosofia, definiu-se o ambiente O-A-V/E-R aqui descrito, que integra sistemas especialistas O-A-V e banco de dados baseados no modelo Entidade-Relacionamento (E-R).

Esta escolha deve-se, entre outros fatores, ao fato de ser o modelo E-R amplamente utilizado na modelagem de dados e de sua facilidade de uso devido à intuitividade de seus conceitos. Além disso, o modelo E-R não fica restrito a implementações particulares de SGBD, podendo ser mapeado para diferentes modelos de banco de dados.

Este trabalho apresenta os fundamentos do ambiente O-A-V/E-R. Inicialmente, são abordados aspectos relacionados ao acoplamento de sistemas especialistas com banco de dados, para que o leitor possa se situar no presente contexto. Como estamos nos referindo ao modelo E-R, dedicamos a seção 3 à introdução dos principais conceitos desta metodologia. A seção 4 apresenta o ambiente O-A-V/E-R, juntamente com um exemplo ilustrando o funcionamento do sistema. Por fim são feitas algumas considerações gerais.

2. ASPECTOS DE ACOPLAMENTO

Num Sistema Especialista (SE) baseado em regras de produção, a máquina de inferência utiliza um conjunto de regras e uma coleção de fatos (Base de Fatos ou Banco de Dados) para simular o comportamento de um especialista humano num determinado problema de domínio específico.

Um banco de dados, por sua vez, é caracterizado em termos de duas dimensões básicas: variedade e cardinalidade. Por exemplo, em representações baseadas em lógica, os diferentes predicados lógicos (ou relações) refletem a variedade, e instâncias destes predicados se referem a cardinalidade.

A maior parte dos BD de SE exibem uma grande variedade de fatos. Por outro lado, a cardinalidade de fatos nestes BD varia de acordo com o problema específico. Assim, os BD de SE diferem dos BD tradicionais no sentido deles serem geralmente mais abrangentes (maior variedade) e menos profundos (menor cardinalidade). Em seguida será apresentada uma abordagem de quatro estágios para o acoplamento de BD à SE.

Quatro estratégias para estabelecer uma comunicação cooperativa entre componentes dedutivos e de dados num Sistema Especialista foram identificadas [JARK 85]. Parte-se das facilidades iniciais e naturais de obtenção de dados , evolui-se para um SGBD generalizado dentro do próprio SE para, então, alcançar o acoplamento fraco de um SE com um SGBD já existente, e finalmente alcança-se o acoplamento forte com um SGBD externo. Projetistas de SE podem optar por uma das estratégias dependendo do volume de dados, multiplicidade de uso dos dados, características voláteis dos dados (quão freqüentemente os dados mudam), e dos requisitos de proteção e segurança. Contudo, num projeto bem delineado estas etapas são incrementais, permitindo transações suaves de um ambiente mais simples para um mais sofisticado.

No *primeiro estágio*, que consiste na manipulação elementar de dados no contexto de um SE, todos os dados são mantidos em memória principal e armazenados em estruturas de dados convencionais. Neste estágio rotinas específicas da aplicação para recuperação e modificação de dados são implementadas.

No *segundo estágio*, que consiste em um SGBD generalizado no escopo de um SE, a quantidade de dados armazenados no BD é grande a tal ponto de não caber na memória principal, neste caso, a manipulação elementar de dados não é suficiente. Algumas técnicas de gerenciamento externo de arquivos são necessárias (índices secundários, dicionário de dados, etc). A implementação destas ferramentas deve ser feita de uma forma modular de modo a ser independente da aplicação.

Outrossim, dependendo da multiplicidade do uso do BD e da extensão do mesmo, facilidades de SGBD podem ser tornar necessárias. Tais facilidades, que incluem vistas e janelas dinâmicas, se encontram disponíveis na maioria dos SGBD relacionais modernos. Outra facilidade consiste em integrar o SE com o dicionário de dados de forma que sejam permitidas consultas a respeito da estrutura do BD. Este estágio de extensão parece ser normal para os SE que lidam com grandes bases de dados, embora nem todos os sistemas os acessem com o mesmo grau de sofisticação. Passa-se, então para o terceiro estágio, denominado acoplamento fraco.

Banco de Dados externos são normalmente robustos, altamente voláteis e utilizados por várias aplicações. Custos de armazenamento e manutenção de consistência proibiriam a duplicação de tais BD para o uso exclusivo do SE.

O acoplamento fraco se refere a presença de um canal de comunicação entre os dois sistemas, pelo qual é feita a extração dos dados do BD existente para posterior armazenagem num BD interno ao SE. A partir deste ponto, pode-se acessar os dados conforme descrito no estágio dois. Observa-se que permite-se apenas operações de consulta.

A principal desvantagem deste tipo de acoplamento consiste no fato dele não se adequar aos casos onde existem decisões dinâmicas. Isto é, durante a mesma sessão de execução de um SE, porções armazenadas em blocos de memória distintos podem ser referenciadas. Deste modo, acessos extras são solicitados gerando um processo de *swapping* que degrada o desempenho final do sistema. Em adição, se o BD externo é continuamente atualizado, a consistência com o interno pode se tornar rapidamente obsoleta.

Assim, surge a necessidade do acoplamento forte entre um SE e um SGBD externo. Neste tipo de acoplamento, o BD externo é visto pelo SE como uma extensão deste. Abordagens deste gênero enfrentam problemas relativos ao número de chamadas ao BD, que em tempo de execução pode prejudicar o desempenho do sistema, ou à complexidade

das consultas, que pode inviabilizar suas traduções para as linguagens de consulta existentes. As considerações mais relevantes na utilização do acoplamento forte consistem no gerenciamento inteligente do canal de comunicação, ou seja, quando e como utilizá-lo, e na possibilidade de se efetuar operações de escrita no BD acoplado.

No acoplamento forte existem duas formas para geração de consultas. A primeira, denominada estática, consiste em utilizar um pré-compilador que realiza a primeira avaliação da lógica do sistema assumindo que todas as referências ao banco de dados serão efetuadas, independente da premissa. Desta forma, o mapeamento dos elementos inerentes ao SE para consultas é feito em tempo de compilação, e na execução as consultas são invocadas diretamente quando a regra em questão for disparada. Na segunda forma de implementação, denominada dinâmica, as consultas são geradas em tempo de execução do sistema.

No contexto deste trabalho adotaremos o acoplamento forte, gerando as consultas de forma estática. Isto se justifica pela possibilidade do banco de dados acessado não estar necessariamente residente no mesmo ambiente do SE e, neste caso, se permitíssemos o acesso dinâmico, estaríamos prejudicando o desempenho final do sistema.

3. INTRODUÇÃO AO MODELO ENTIDADE-RELACIONAMENTO

Nesta seção são apresentados os principais elementos que definem o Modelo Entidade-Relacionamento. Para o leitor mais interessado no assunto recomenda-se a leitura de outros trabalhos aqui referenciados [SILV 89].

Os elementos básicos para a definição de um esquema E-R são:

- a) *Constantes*, que são denotadas por c_1, c_2, \dots ; e ϕ para o valor nulo.
- b) *Conjuntos de Entidades e Relacionamentos* que são identificados como E_1, E_2, \dots e R_1, R_2, \dots , respectivamente. Cada relacionamento possui associado um inteiro $dg(R)$, dito ser o grau de R , que consiste no número de entidades ligadas por R . Deste modo, $dg(R) \geq 2$, e $E_1, \dots, E_{dg(R)}$ é a lista das entidades associadas por R , ou lista das entidades participantes.
- c) Para cada relacionamento R existe um conjunto de funções papéis da forma

$$\{\rho_1 : R \rightarrow E_1, \dots, \rho_{dg(R)} : R \rightarrow E_{dg(R)}\}$$

onde cada E_i é uma entidade participante de R . Funções papéis são utilizadas para identificar instâncias das entidades ligadas por um relacionamento R .

- d) *Conjuntos de Valores*, que são denotados por V_1, V_2, \dots e são utilizados para definir

conjuntos de constantes, especificando os domínios dos atributos. A cada conjunto V está associado um conjunto de constantes $\{c_1, \dots, c_n\}$.

e) Símbolos de atributos, α_1, α_2 que são funções do tipo

$$\alpha : X, I \rightarrow V$$

onde X é um conjunto de entidades ou relacionamento, I é um inteiro, e V é um conjunto de valores. Cada atributo α possui associado um inteiro $maxoc(\alpha)$, $1 \leq I \leq maxoc(\alpha)$. Na realidade, este inteiro especifica o número de valores associado ao atributo. Assim, para atributos monovalorados este valor é 1 e ≥ 1 para os multivalorados.

f) *Restrições de Integridade* são fórmulas bem formadas $\Theta_1, \Theta_2, \dots$. Restrições de integridade estão fora do escopo deste trabalho. Maiores detalhes a respeito desta fórmulas podem ser obtidos nos artigos aqui referenciados [SILV 89].

Com respeito ao item f , podemos considerar alguns elementos do modelo E-R como restrições de integridade, entre os quais destacamos:

- Chaves
- Cardinalidades de relacionamento
- Entidades Fracas

No contexto deste trabalho apenas o primeiro tópico será abordado.

Chaves

Chaves determinam conjuntos de atributos que identificam unicamente instâncias de um dado conjunto de entidades ou relacionamento. De forma a garantir a implementação correta das chaves, é necessário garantir a propriedade de unicidade. Para isso são definidas funções auxiliares que manipulam chaves [SILV 90]. É importante ressaltar que no ambiente O-A-V/E-R as chaves serão um dos meios de acesso ao banco de dados.

4. O AMBIENTE O-A-V/E-R

Uma vez apresentados os principais elementos do modelo E-R, podemos definir o ambiente O-A-V/E-R. Este novo ambiente consiste da integração dos elementos E-R com os definidos para bases de conhecimento baseadas em triplas O-A-V e regras de produção [HARM 85], como será descrito nesta seção.

Triplas O-A-V

Triplas O-A-V são utilizadas para representação de conhecimento em base de fatos e, neste trabalho, a seguinte notação será utilizada: 1

Objetos. $o, o_1, o_2,$

Atributos. $\alpha, \alpha_1, \alpha_2,$

Valores. $c, c_1, c_2,$ e ϕ para o valor nulo.

A semântica de (o, α, c) é definida como: o atributo α de um objeto o assume o valor c . De outra forma, podemos enxergar os atributos como se fossem funções, que aplicadas a objetos retornam valores. Assim, pode-se considerar a atribuição $\alpha(o) = c$ equivalente à existência da tripla (o, α, c) . Podemos, então, estabelecer a seguinte definição:

Definição 1. *Uma base de fatos B é constituída por um conjunto de triplas O-A-V da forma (o, α, c) .*

Foi observado que esta definição poderia ser estendida para os dados armazenados em um banco de dados, e, desta forma, a definição postulada acima pode ser interpretada como: uma base de dados é constituída por um conjunto de triplas O-A-V (o, α, c) , onde o se refere ao conjunto de entidades ou relacionamentos em questão, α corresponde ao atributo a ser lido ou atualizado, e c consiste no valor do atributo α . Do mesmo modo, a semântica da expressão $\alpha(o) = c$, neste contexto é: o atributo α do objeto o assume o valor c . o representa um elemento de um conjunto entidades ou relacionamentos X , que é obtida através da função F , a ser descrita no próximo tópico.

Percebeu-se, então, que se fossem redefinidos os termos, literais e regras de modo a introduzir os elementos E-R no escopo O-A-V, e acrescentado mecanismos de quantificação existencial e universal, poderíamos obter o almejado ambiente O-A-V/E-R. A descrição desses novos elementos aparece a seguir.

Termos O-A-V/E-R

Os termos O-A-V/E-R podem ser considerados como canal de comunicação entre a base de conhecimento e o banco de dados. Deste modo, os termos O-A-V/E-R podem ser sub-divididos em três tipos:

1 Na realidade, há pequeno abuso de notação aqui, uma vez que a notação para atributos, objetos e valores é a mesma utilizada na seção 3. A interpretação correta, entretanto, é facilmente diferenciada pelo contexto.

Termos-O

Denotam objetos da base de fatos B , que podem ser:

- a) um objeto o de B
- b) uma variável-O, que tem como domínio um conjunto de objetos de B

Termos-E

Denotam elementos das instâncias de conjuntos de entidades e relacionamentos, que podem ser:

- c) um objeto abstrato o associado a um elemento específico de um conjunto de entidades ou relacionamentos
- d) uma variável-E-R, que tem como domínio um conjunto de entidades ou relacionamentos
- e) uma função papel $\rho(t)$, onde t é um Termo-E
- f) uma função chave $F(X, t_1, \dots, t_n)$, onde t_i é cada um dos componentes de alguma chave primária de X . Na realidade esta função retorna um objeto abstrato o pertencente à instância de X

Termos-C

Denotam constantes, que podem ser:

- g) uma expressão da forma $\alpha(t)$ onde α simboliza um atributo de t , onde t é um Termo-O
- h) um símbolo de função de grau n seguido por uma lista da forma (t_1, \dots, t_n) , onde t_i é um termo-C
- i) uma constante c
- j) uma expressão da forma $\alpha(t)$, onde α é um atributo de t , onde t é um termo-E ²
- k) uma variável-C que tem como domínio um conjunto de constantes

Literais

Os literais, no ambiente O-A-V/E-R, permitem o uso de operadores relacionais, comum a ambientes O-A-V, e testes de pertinência. Testes de pertinência são necessários para verificar a existência de algum elemento num conjunto de entidades ou relacionamentos ou, no caso de atributos multivalorados, a pertinência de um valor qualquer na

² Os itens g e j , estritamente falando, são da forma $\alpha(t, i)$, onde i referência o i -ésimo valor da lista de valores de um atributo multivalorado

listas de valores definidos para o atributo em questão. Procurou-se, também, definir os literais no mais alto nível possível de modo a tornar transparente para o usuário as operações específicas do banco de dados [SILV 90a]. Assim foram definidos as seguintes formas de literais:

- a) $t_1 \theta t_2$, onde θ representa um operador relacional
- b) $t \in X$, onde X representa um conjunto de entidades ou relacionamentos
- c) $t_1 \in \alpha(t_2)$, onde α representa um atributo multivalorado

Para os três literais definidos acima, t_i é um termo O-A-V/E-R.

Regras

As regras são compostas de literais e ações e possuem a seguinte notação:

Regras. R, R_1, \dots, R_n .

Ações. A, A_1, \dots, A_n .

Literais. L, L_1, \dots, L_n .

Uma regra R é definida como

$$v_1 \in D_1, \dots, v_n \in D_n$$

$$A(v_1, \dots, v_n) \leftarrow L_1(v_1, \dots, v_n), \dots, L_k(v_1, \dots, v_n),$$

$k \geq 0$ e D_i representa o domínio da variável v_i , que pode ser um conjunto de entidades ou relacionamentos, um conjunto de objetos de B ou um conjunto de constantes.

Admite-se o uso de variáveis livres desde que toda variável v_i que esteja numa literal apareça, obrigatoriamente, na ação da regra. Esta restrição se faz necessária para viabilizar o processo de inferência.

Normalmente, uma ação provoca a execução de algum procedimento e o que nos interessa, no escopo do ambiente O-A-V/E-R, é a inclusão de uma nova tripla (o, α, c) na base de fatos B , expressa por:

$$t_1 \leftarrow t_2$$

onde t_1 e t_2 podem ser qualquer um dos termos O-A-V/E-R definidos acima.

Avaliação de Regras e Literais

Neste trabalho assumimos que os termos não possuem fatores de certeza. Assim,

para a avaliação de um literal foi definida uma função v que aplicada ao seu argumento retorna um valor no domínio $\{0, 1\}$, correspondendo à avaliação verdadeira ou falsa, respectivamente, de seu argumento. Assim, podemos postular duas definições para a avaliação de literais e regras.

Definição 2. *Seja L uma literal da forma $t_1 \theta t_2$. A avaliação de L é:*

$$v(L) = \begin{cases} 0, & \text{se } t_1 \theta t_2 \text{ é falso;} \\ 1, & \text{se } t_1 \theta t_2 \text{ é verdade;} \\ \text{indefinido,} & \text{se } t_1 \text{ e } t_2 \text{ forem incompatíveis.} \end{cases}$$

Definição 3. *Seja L uma literal da forma $t \in X$. A avaliação de L é:*

$$v(L) = \begin{cases} 0, & \text{se } t \text{ não é um elemento de } X; \\ 1, & \text{se } t \text{ é um elemento de } X; \\ \text{indefinido,} & \text{se } t \text{ e } X \text{ forem incompatíveis.} \end{cases}$$

Definição 4. *Seja L uma literal da forma $t_1 \in \alpha(t_2)$, onde α é um atributo multivalorado. A avaliação de L é:*

$$v(L) = \begin{cases} 0, & \text{se } t_1 \text{ não está contido na lista de valores de } \alpha(t_2); \\ 1, & \text{se } t_1 \text{ está contido na lista de valores de } \alpha(t_2); \\ \text{indefinido,} & \text{se } t_1 \text{ e } \alpha(t_2) \text{ forem incompatíveis.} \end{cases}$$

Definição 5. *Seja R uma regra da forma*

$$v_1 \in D_1, \dots, v_n \in D_n$$

$$A(v_1, \dots, v_n) \leftarrow L_1(v_1, \dots, v_n), \dots, L_k(v_1, \dots, v_n),$$

$k \geq 0$ e D_i representa o domínio da variável v_i . A avaliação de R é:

$$v(R) = \begin{cases} 0, & \text{se } v(L_i) = 0, \text{ para algum } i \in [1, k]; \\ 1, & \text{se } v(L_i) = 1, \text{ para todo } i \in [1, k] \end{cases}$$

Na avaliação de uma regra R , deve existir uma substituição $\nabla = [t_1/v_1, \dots, t_n/v_n]$, tal que todas as ocorrências de v_i em R devem ser substituídas pelo termo t_i correspondente em ∇ .

Para ilustrar a utilização de elementos O-A-V/E-R, considere o esquema E-R da Figura 1, onde são definidos os seguintes componentes:

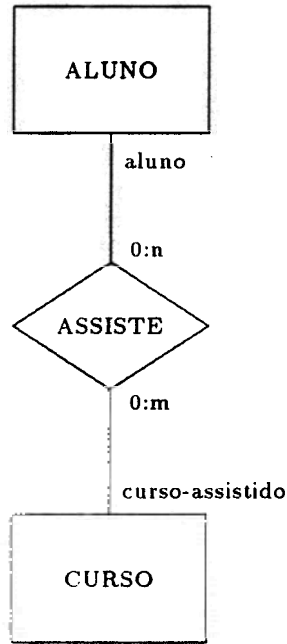


Figura 1. Um exemplo simplificado de um diagrama E-R

Entidades:

ALUNO

CURSO

Relacionamento:

ASSISTE

Atributos:

código(CURSO) → inteiro

nome(CURSO) → alfanumérico

periodo(CURSO) → { manhã, tarde, noite }

faculdade(CURSO) → alfanumérico

nome(ALUNO) → alfanumerico

idade(ALUNO) → inteiro

Funções Papel:

curso-assistido(ASSISTE) → CURSO

aluno(ASSISTE) → ALUNO

Para este esquema, as seguintes regras podem ser definidas:

Regra 1

$c \in \text{CURSO}$

Se $\text{nome}(c) = \text{"Matemática"}$

e $\text{faculdade}(c) \neq \text{"Ciência Exatas"}$

então $\text{periodo}(c) = \text{"tarde"}$

Regra 2

$\text{cur} \in \text{CURSO}$

Para algum $\text{curso} \in \text{CURSO}$

Se $\text{periodo}(\text{curso}) = \text{periodo}(\text{cur})$

e $\text{nome}(\text{cur}) \neq \text{"FISICA"}$

então $\text{código}(\text{curso}) = f(\text{cur})$

Regra 3

Para algum $\text{aluno} \in \text{ALUNO}$

Se $\text{idade}(\text{aluno}) \leq \text{idade}(F(\text{ALUNO}, \text{"Maria"}))$

e $\text{curso-assistido}(\text{aluno}) = \text{"História"}$

então $\text{ocorrência} = \text{TRUE}$

Regra 4

Se $\text{faculdade}(\text{CURSO}) = \text{"Ciências Humanas"}$

então $\text{periodo}(\text{CURSO}) = \text{"Noite"}$

Neste exemplo podemos identificar os seguintes termos:

- $\text{nome}(c)$, na regra 1, conforme definido no item j
- aluno , na regra 3, conforme definido no item d
- ocorrência , na regra 3, conforme definido no item k
- $\text{curso-assistido}(\text{aluno})$, na regra 3, conforme definido no item e
- $f(\text{cur})$, na regra 2, conforme definido no item h
- $F(\text{ALUNO}, \text{"Maria"})$, na regra 3, conforme definido no item f

Instanciações Existenciais

Observa-se, nas regras 2 e 3, a presença do quantificador existencial "para algum". Nestes dois casos, a instanciação se dá através de uma busca nas tabelas referentes a CURSO e ASSISTE, respectivamente. Como não foi pré-estabelecida uma ordem de busca, o SGBD pode utilizar o critério que lhe convier, que em muitos casos pode afetar o desempenho final do sistema. Assim, como evolução deste trabalho, está prevista a introdução do termo ORD, que possibilita a busca ordenada como ilustrado a seguir:

Para algum curso \in CURSO através de Curso.Alfabética

Esta expressão estabelece a instanciação de *curso* através da busca por ordem alfabética no conjunto de entidades CURSO. Note-se que a definição original do Modelo Entidade-Relacionamento não faz referência às ordens de acesso, uma vez que sua utilização inicialmente pretendida era apenas como uma simbologia para a descrição lógica de um sistema de banco de dados. Com a utilização aqui empregada, foi necessário introduzir o conceito de ordem nos diagramas E-R [SILV 90b].

O escopo das variáveis instanciadas existencialmente abrange apenas a regra onde houve sua instanciação. Assim, no exemplo, o escopo de *aluno* é a regra 3.

Instanciações Universais

Estabeleceu-se que as variáveis livres serão instanciadas pelo usuário na inicialização do sistema e atingem as regras onde elas aparecem. O mecanismo de avaliação das regras segue o algoritmo de *backtracking* [HARM 85]. Todas as variáveis livres que aparecem na conclusão deverão ser instanciadas no momento que a regra for disparada, de acordo com o critério que estabelece que toda variável v_i que esteja numa literal apareça obrigatoriamente, na ação da regra.

5. CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma proposta para a integração de sistemas especialistas e bancos de dados, tomando-se por base o paradigma O-A-V para representação do conhecimento. Embora o assunto seja extenso e apresente numerosas questões a serem resolvidas, parece-nos que a abordagem ataca os principais problemas de tais interfaces e estabelece a forma geral para a referida integração. A utilização de termos que contemplam elementos de bancos de dados baseados no Modelo E-R é uma solução robusta, completa e primordialmente simples e intuitiva, não trazendo perturbações ao modelo originalmente empregado para a construção de regras de produção.

Há várias extensões a considerar num trabalho desta natureza, entre as quais podemos citar as seguintes.

Atualizações da base de dados

Bancos de dados são normalmente orientados para o processamento de transações, que são caracterizadas por unidades de processamento. Entre o início de uma transação e seu final, o banco de dados pode passar por estados inconsistentes, embora seu estado resultante deva ser um estado consistente. Uma transação também caracteriza os elementos para a recuperação de um banco de dados por ocasião de falhas, cancelando a transação ou executando-a completamente. Se considerarmos que as ações resultantes de uma regra podem gerar alterações na base de dados, será necessário considerar a ativação de procedimentos de *begin-transaction* e *end-transaction*, de modo a comunicar à base de dados os limites de consistência das operações e instruí-lo quanto ao procedimento de recuperação. Embora este problema resulte em algumas pequenas alterações na codificação das regras aqui introduzidas, ele não compromete a solução acima delineada. A solução aqui descrita suporta atualizações e demarcação de transações no presente ambiente de integração.

Linguagem para a codificação das regras

O presente artigo não sugere uma sintaxe definitiva para as regras que podem ser escritas com os termos estendidos. Esta linguagem será melhor definida no decorrer da implementação, embora todos os seus elementos sintáticos tenham sido aqui apresentados. O que resta é apenas encontrar a melhor forma de modo a assegurar perturbações mínimas para implementações específicas, ou seja, manter o máximo de semelhança com a sintaxe suportada pela *shell* à qual o ambiente de integração será agregado.

Implementação experimental

Está em curso uma implementação experimental do presente ambiente de integração. Para tanto, utiliza-se uma *shell* previamente implementada, de nome BACO II [PACI 90], sem nenhuma comunicação com bases de dados externas, embora a mesma dispusesse de uma interface genérica para chamadas às funções externas. A referida *shell* já oferece um forma para tratamento de incertezas que será alterada face à introdução dos termos oriundos da base de dados externa.

Bancos de Dados E-R

Bancos de Dados E-R não são comumente encontrados em implementações comerciais, embora haja uma tendência atual neste sentido. O ambiente de integração aqui

mencionado faz parte de um projeto de maior porte, chamado Projeto UniversiData, onde se procura estabelecer uma ambiente completo para bancos de dados E-R.[SILV 90a] Uma das interfaces disponíveis para os usuários é uma *shell* para a construção de sistemas especialistas com acesso aos dados armazenados. As funções de bancos de dados por ela utilizadas foram especialmente definidas para prover uma interface externa consistente e de uso geral.

6. REFERÊNCIAS

- [CHEN 76] CHEN, P.: The Entity-relationship Model - Toward a Unified View of Data, *ACM Transactions on Database Systems* 1, (1976).
- [CLAN 87] CLANCEY, W.J.: *Knowledge-Based Tutoring: The GUIDON Program*, MIT Press, Cambridge, Massachusetts, 1987.
- [FORS 84] FORSYTH, R.: *Expert Systems: Principles and case studies*, Chapman and Hill, 1984, USA.
- [GARD 89] GARDARIN G., VALDURIEZ P.: *Relational Databases and Knowledge Bases*, Addison-Wesley Publishing Company, 1989.
- [HARM 85] HARMON, P. and KING, D.: *Expert Systems*, John Wiley & Sons, Inc., USA, 1985.
- [HORN 86] HORN, N.V.: *Understanding Expert Systems*, Bantam Books Inc., New York, 1986.
- [JARK 85] JARKE, MATTHIAS, VASSILIOU, YANNIS: Coupling Expert Systems with database Management Systems. In: Reitman, W. (ed). *Artificial Intelligence Applications for Business*, Norwood, NJ:Ablex, 1985, cap. 5 p. 65-85
- [MANN 74] MANNA, Z.: *Mathematical Theory of Computation*, McGraw-Hill Computer Science Series, 1974.
- [OSHE 84] O'SHEA, T. e M. EISENSTADT: *Artificial Intelligence, Tools, Techniques and applications*, Harper and Row, New York, 1984.
- [PACI 90] PACITTI, E.C., MENDES F.F., TELES A.S.: *Um Gerenciador de Bases de Conhecimento Multivaloradas*, a ser publicado como Relatório Técnico NCE/UFRJ.
- [PECK 88] PECKHAM, M. and MARYANSKI, F.: Semantic Data Models, *ACM Computing Surveys* 20, 3(September 1988).
- [RICH 83] RICH, E.: *Artificial Intelligence*, McGraw-Hill, 1983, USA.

- [SETZ 86] SETZER, W.: *Projeto Lógico e Projeto Físico de Bancos de Dados*, V Escola de Computação, Belo Horizonte, 1986.
- [SILV 89] SILVEIRA, P. M.: A Formalization of The E-R Model, *Proceedings of the IX Conferencia Internacional de la Sociedad Chilena de la Ciencia de la Computation*, Santiago, Chile, 1989.
- [SILV 90a] SILVEIRA, P. M.: Definindo e Utilizando Bancos de Dados com o Modelo Entidade-Relacionamento, in *Anais do XXIII Congresso Nacional de Informática*, Rio de Janeiro, 1990.
- [SILV 90b] SILVEIRA, P. M.: Procedural Data Manipulation Operations for the E-R Model, *Proceedings of the XVI Conferência Latino Americana de Informática*, Assunção, Paraguay, 1990.
- [SILV 86] SILVEIRA, P. M., Aspectos da Integração de Banco de Dados e Sistemas Especialistas, *Anais do II Congresso Informazônia*, São Luis, 1986.