

FERRAMENTA DE SOFTWARE PARA O DESENVOLVIMENTO DO CICLO
DE VIDA DE UM PROJETO EM BANCO DE DADOS A PARTIR DE TÉCNICAS
DE INTELIGENCIA ARTIFICIAL

Francisco Gonçalves Pereira Neto
Odair Fernandes Aguiar Filho
Ana Maria de Carvalho Moura

Instituto Militar de Engenharia
Departamento de Informática
Pça. Gen. Tibúrcio 89 - Urca
22290 Rio de Janeiro
Tel: (021) 295-32-32 - Ramal 214

SUMARIO

O entendimento da realidade, através de uma descrição clara e precisa, veio dar origem ao estudo da representação do conhecimento humano, objeto de pesquisa dos mais diferentes ramos da ciência. Dentro da Informática três áreas tem se destacado: Inteligência Artificial (IA), Banco de Dados (BD) e Linguagens de Programação (LP). Este trabalho enfocará os esforços de pesquisa dentro das áreas de BD e IA, principalmente aqueles que visam as aplicações de técnicas de IA na modelagem conceitual de BD.

1- INTRODUÇÃO

Uma das áreas da informática que têm merecido mais atenção nos últimos anos é a de projeto de sistemas de informação.

Apesar de todos os avanços teóricos obtidos nesta área, pouca utilização de tais conhecimentos têm ocorrido a nível prático e aspectos tais como normalização até a 5.ª forma normal (5 FN) em banco de dados relacionais e restrições de integridade referenciais ainda soam apenas como teoria acadêmica.

Dentre os fatores que surgem como maiores obstáculos para que tais avanços passem da teoria à prática, podemos citar:

- dificuldade dos usuários na absorção de tais técnicas, tendo em vista a utilização de instrumentos de formalização sofisticados tais como redes semânticas, cálculo de predicados, redes de Petri, etc., além da simbologia e conceitos próprios a cada metodologia;

- o nível de complexidade para a especificação de aplicações de grande porte aumenta consideravelmente, mesmo empregando metodologias sofisticadas;

- a preservação das características especificadas para um sistema após a realização de um conjunto qualquer de operações sobre os seus dados não é trivial, sendo hoje em dia objeto importante de pesquisas.

A solução proposta para a minimização destes obstáculos é a construção de ferramentas automatizadas, que reúnem o conhecimento teórico embutido nas áreas de BD e IA.

Dentro desta linha de pesquisa surgem sistemas especialistas para projeto de sistemas de informação [BOU86]. Tais sistemas teriam como característica adicional a capacidade de deduzir conhecimentos não explicitamente declarados pelo usuário bem como explicar os passos levados em conta para a elaboração do projeto.

Este trabalho propõe o desenvolvimento de uma ferramenta de software que permita a automatização de todo o ciclo de vida de projeto de um BD, desde o nível externo (visões) até seu nível interno, passando pelas fases de modelagem conceitual e lógica de dados. Para isso, foi necessário integrar ao conhecimento geral de metodologias de projeto de BD, ferramentas básicas utilizadas dentro da IA para representação do conhecimento, tais como Lógica e Rêdes Semânticas, de modo a construir uma ferramenta de auxílio a um projetista de BD, provendo-a de uma interface amigável que minimize o conhecimento necessário a sua utilização.

2- O CONHECIMENTO HUMANO E SUA REPRESENTAÇÃO

2.1- REPRESENTAÇÃO DO CONHECIMENTO ATRAVES DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

A representação do conhecimento é um problema central da Inteligência Artificial (IA) nos dias de hoje [BRO84]. Sua resolução depende do desenvolvimento de uma notação que permita representá-lo com precisão. Mylopoulos sugere a seguinte classificação para as formas de representação de conhecimento: Redes semânticas, Lógica, Regras de Produção e "Frames". Cada uma dessas abordagens apresenta uma forma de representação própria (procedural ou declarativa), apesar de possuírem características globais em comum.

2.2- REPRESENTAÇÃO DO CONHECIMENTO ATRAVES DE TÉCNICAS DE BANCO DE DADOS (MODELOS DE DADOS)

Um modelo de dados [TSI82] tem por objetivo definir regras gerais para a especificação das estruturas de dados e das operações permitidas sobre os mesmos. Deve ainda tentar capturar as propriedades estáticas e dinâmicas da fatia da realidade a ser modelada. As propriedades estáticas correspondem àquelas que são relativamente invariantes com o tempo, tais como: objetos, propriedades e relacionamentos entre objetos. As propriedades dinâmicas correspondem a evolução natural do mundo, tais como: operações sobre

objetos, propriedades das operações e relacionamentos entre as mesmas. As propriedades dinâmicas são definidas como funções parciais sobre o estado da base de dados. A função é dita parcial porque pode mapear uma ocorrência da base de dados em uma outra de estrutura bem definida, ao mesmo tempo em que viola algumas restrições associadas com o modelo. Qualquer modelo de dados deve ser de alguma maneira capaz de capturar estas duas propriedades.

Para situar o estado atual da arte na modelagem conceitual de dados podemos classificá-la segundo as metodologias descritas em [BR084]:

- Modelos de dados primitivos.
- Modelos de dados clássicos: Hierárquico, Redes, Relacional
- Modelos de dados semânticos:
 - Extensões dos modelos de dados clássicos: ER [CHE76].
 - Modelos de dados matemáticos [CHI77].
 - Modelos de dados irredutíveis [HAL76].
 - Modelos de dados semânticos hierárquicos estáticos:
 - SHM [SMI77], RM/T [COD79], SDM [HAM81], MORSE [BOU86].
 - Modelos de dados semânticos hierárquicos dinâmicos:
 - SHM+ [BR081], TAXIS [MYL78].
- Modelos de dados para propósitos especiais.

3- PROJETO DE UMA FERRAMENTA DE SOFTWARE PARA CONCEPÇÃO DE BANCO DE DADOS

O objetivo deste projeto, é construir uma ferramenta dirigida ao Analista/Projetista de Banco de Dados e tem por finalidade auxiliá-lo em todas as etapas de concepção, desde a análise dos requisitos e especificação das visões externas até a estrutura interna, passando pela integração de visões, normalização de relações e otimização das estruturas do nível interno. Tal ferramenta será constituída de cinco módulos, cada um dos quais envolvendo uma fase do projeto de banco de dados. A opção por vários módulos deve-se a complexidade do problema (projeto de BD) e a possibilidade deste ser dividido em partes bem definidas (fases de projeto ou concepção). A escolha da tecnologia de SE (Sistemas Especialistas) para a sua implementação deve-se as seguintes razões:

possibilidade de técnicos pouco experientes poderem substituir especialistas largamente experimentados;

- possibilidade de utilização do sistema para a formação de novos especialistas;

- o conhecimento obtido do especialista fica registrado, podendo ser facilmente difundido;

- a utilização do sistema induz o usuário a organizar e sistematizar a tarefa a ser executada.

Os níveis de abstração, as diferentes fases de concepção, bem como os módulos constituintes do sistema são mostrados na figura 1.

4.1- AQUISIÇÃO DE CONHECIMENTO

A interação Homem-Máquina vem despertando interesse nos últimos anos como uma decorrência da evolução dos sistemas de computação. Alguns motivos que tem levado a uma maior preocupação com estes aspectos são:

- O crescimento da complexidade destes sistemas;
- A popularização do computador com a sua introdução em diversas áreas da atividade humana.

O sucesso da maioria dos sistemas de computação depende diretamente da sua habilidade de interagir com o usuário. Parece ser de aceitação universal que esta interação deveria ser feita através de linguagem natural, não obrigando o usuário a aprender uma linguagem artificial menos expressiva. No entanto a utilização de um sistema de linguagem natural requer uma sobrecarga de processamento bastante acentuada. Além disso, o atual estágio de desenvolvimento destes sistemas permite a utilização de apenas um subconjunto da linguagem, restringindo mais uma vez o vocabulário do usuário àquele utilizado pelo sistema.

Nossa maior preocupação durante o desenvolvimento da interface para aquisição de conhecimento, foi a de tornar cada vez mais transparente ao usuário o fato de que todo o conhecimento por ele transmitido seria representado internamente através de nós e arcos de uma rede semântica.

As vantagens apresentadas pelas Redes Semânticas são: capacidade de representação de uma grande variedade de tipos de objetos, simplicidade e facilidade de representação do conhecimento através desta estrutura. Dentro desta concepção utilizamos a rede semântica MORSE [BOU86], e [MPN87] que procura aproveitar as melhores características dos modelos semânticos ER, Binários [ABR74], SHM, RM/T, SDM, SHM+. Uma rede semântica pode ser vista como um grafo orientado, constituído de um conjunto de nós e de arcos interligando-os entre si. A rede semântica proposta é definida pela tripla $RS\langle CN, CA, CR \rangle$, onde: CN - Categoria de nós, CA - Categoria de Arcos e CR - Categoria de Restrições de Integridade.

4.1.1- ENFOCANDO OS NÓS DA REDE SEMÂNTICA COMO OBJETOS

São quatro as categorias dos nós (CN) da rede semântica RS: Atributos, Entidades, Valores de atributos, Instâncias de Entidades.

(i) Atributo: Dentro da RS proposta, é um conceito atômico com relação a sua estrutura, podendo ser propriedade de uma ou diversas entidades.

(ii) Entidade: É um conceito molecular, uma vez que a mesma pode ser constituída da agregação de vários atributos ou mesmo de outras entidades.

(iii) Valor de atributo: É um conceito atômico, sendo a

menor informação significativa manipulável. Os valores de um atributo podem ser inteiros, reais, strings e booleanos.

(iv) Instância de entidade: é a agregação de valores dos atributos descrevendo um ou diversos fatos da realidade considerada.

Smith & Smith [SMI77] enfatizam o aspecto "relativismo semântico" não suportado por alguns modelos de dados, mais especificamente quanto ao fato destes obrigarem o utilizador a declarar o objeto e o seu tipo, não permitindo que outros o vejam de forma diferenciada, ou mesmo que o tipo de um objeto seja declarado desconhecido. Tais limitações foram ultrapassadas na atual implementação onde o modelo utilizado suporta não só múltiplas visões de um mesmo objeto como também o desconhecimento da categoria onde se enquadram, ficando a cargo do sistema a dedução do seu tipo e dos relacionamentos que este mantém com outros objetos.

Algumas das mais avançadas pesquisas dentro da área de BD dizem respeito aos bancos de dados orientados para objetos, cujos requisitos principais citados em [ZAN 86] são os seguintes: abstração e encapsulamento de dados, identidade dos objetos independente dos valores de suas propriedades, herança de propriedades, interfaces interativas através de janelas e menus, etc, atendidos parcialmente na versão atual da ferramenta.

4.1.2- REPRESENTAÇÃO DOS ARCOS DA REDE SEMÂNTICA ATRAVÉS DE EXPRESSÕES EM LINGUAGEM NATURAL

Cada nó X da RS pode ser ligado a outros nós por um conjunto de arcos, de acordo com as categorias destes nós [MNP87].

- Arco de agregação de átomos $a(X,Y)$ e particularização $p(Y,X)$,
- Arco de agregação de moléculas $am(X,Y)$ e participação $pm(Y,X)$,
- Arco de classificação de objetos $c(X,Y)$ e instanciação de objetos $i(Y,X)$,
- Arco de generalização $g(X,Y)$ e especialização $es(Y,X)$,
- Arco de equivalência $e(X,X')$.

A tabela abaixo mostra a correspondência entre nós e arcos da rede semântica MORSE, utilizada na ferramenta.

!NOS	! ENTIDADE	! ATRIBUTO	! INSTANCIA	! VALOR	!
!ENTIDADE	!am/pm,g/es!	a/p	i/c	-	!
!ATRIBUTO	p/a	-	-	i/c	!
!INSTANCIA	c/i	-	am/pm	p/a	!
!VALOR	-	c/i	a/p	-	!

A versão atual do sistema suporta uma interface totalmente amigável, e a interação com o usuário é realizada através de janelas, menus, ícones, perguntas x respostas, e exemplos.

4.1.3- CATEGORIAS DE RESTRIÇÃO DE INTEGRIDADE (CR)

Restrições de Integridade são regras usadas para definir as propriedades estáticas e dinâmicas que não foram convenientemente expressas pelas características próprias aos objetos e operações do modelo.

(i) Restrição de domínio: Esta restrição é representada por um arco $D(at,d)$ onde at representa o atributo e d o domínio.

(ii) Restrição de especialização: A cada arco "es" pode ser associada uma restrição sob a forma de uma expressão predicativa.

(iii) Restrição de cardinalidade: As cardinalidades são restrições associadas aos arcos a/p e am/pm. Estas cardinalidades são representadas por um par de valores (m,n) que especificam se a relação é total ($m=1$) ou parcial ($m=0$) e também se a relação é funcional ($n=1$) ou não ($n>1$).

(iv) Restrição de dependência:

(iv.i) Dependência funcional: Esta restrição é a mesma que aquela definida dentro do modelo relacional, quando aplicada aos atributos de um objeto. É expressa através de um arco $df(X,Y)$ ligando dois atributos X,Y , sendo Y dependente funcional de X .

(iv.ii) Dependência multivalorada: De modo semelhante à dependência funcional, ela é expressa através de um arco $dm(X,Y)$.

(v) Restrição de unicidade: A cada entidade pode ser associada uma restrição de unicidade sobre um atributo ou coleção de atributos desta entidade.

(vi) Restrição de função: Esta restrição é específica para os arcos de agregação molecular (am), sendo um complemento indispensável para o conhecimento da semântica do arco am. Seu objetivo é definir precisamente a função dos componentes na associação resultante.

(vii) Restrição de união e interseção: Esta restrição é específica a um conjunto de nós e arcos que constituem uma hierarquia de generalização. Se X é a generalização de Y ($g(X,Y)$), X herdará todas as propriedades de Y (restrição de união). A interseção entre duas classes X_1 e X_2 é possível quando existir uma outra classe Z tal que $g(X_1,Z)$ e $g(X_2,Z)$.

4.2- DEDUÇÃO

Uma das características mais importantes da ferramenta é o fato dela aceitar o conhecimento incompleto e através de regras de dedução tentar complementar a descrição feita pelo usuário. Algumas vezes, no entanto, a tentativa de dedução pode fracassar, o que fará com que o sistema recorra ao pedido de exemplos, com os quais tentará elucidar suas incertezas. Se estas permanecerem, tentará induzir o usuário a um pedido de auxílio ou mesmo a recorrer ao módulo de ensino para que eventuais dúvidas sejam sanadas.

5- SISTEMA DE BANCO DE DADOS INTELIGENTE

Em [MAP87] é proposto um sistema que utilize a integração de Programação Lógica e Teoria de Banco de Dados Relacionais para a concepção e utilização de um Sistema de Banco de Dados Inteligente(SBDI), que tenha as seguintes características:

(i) conhecimento de uma metodologia sofisticada de projeto de sistemas de informação que permita a sua utilização por usuários com um mínimo de conhecimento através de interface amigável provendo a capacidade de especificar aplicações de grande porte de maneira consistente;

(ii) manutenção das características especificadas no projeto, através de procedimentos que mantenham ativo o conhecimento obtido durante as operações sobre a base de dados;

(iii) capacidade dedutiva, i.e.,possibilidade de, a partir, das informações explícitas(fatos elementares e regras gerais) obter novas informações através de consultas ou de relações virtuais.

O SBDI compõe-se de 3 sistemas com funções específicas: um para modelagem conceitual e geração do esquema lógico, outro para acesso a base de dados que mantém ativo o conhecimento obtido no primeiro e que permite o acesso a base de maneira dedutiva através da inclusão de relações virtuais, e finalmente um sistema de gerência de banco de dados.

6- CONCLUSÃO

No IME as pesquisas sobre integração de BD e IA são recentes e estão sendo levadas a termo em duas linhas de pesquisa, uma voltada para utilização de técnicas de IA no projeto de BD e outra para pesquisas sobre bancos de dados dedutivos. Pretende-se, a curto prazo, a integração destas duas linhas, com a construção de uma ferramenta de software para modelagem conceitual de dados e um banco de dados com poder dedutivo.

7- BIBLIOGRAFIA:

- [BOU86] Bouzeghoub, M. SECSI - Un Système Expert en Conception de Systèmes d'Informations, Modélisation Conceptuelle de Schémas de Bases de Données. Dissertação Doutorado. L'Université Paris VI. 1986.
- [BR081] Brodie, M.L. Association: A Database Abstraction for Semantic Modelling. Proc. 2nd International Entity-Relationship Conference, 1981.
- [BR084] Brodie, M.L. et alii. On Conceptual Modeling. Springer Verlag, 19-47, 1984.

- [CHE76] Chen, P.P. Entity-Relationship Model - Toward a Unified View of Data - ACM TODS, vol 1, no. 1, pp. 9-36. 1976.
- [CHI77] Childs, D.L. Extended Set Theory. Proc. 3rd International Conference on Very Large Databases.
- [COD79] Codd, E.F. Extending the Database Relational Model to Capture More Meaning. ACM Transaction on Database System, Vol.4, no. 4, pp. 397-434. 1979. Tokyo. Japan. 1977.
- [HAL77] Hall, P., et alii. Relations and Entities. Modelling in Database Management Systems, Elsevier North-Holland, New York, 1976.
- [HAM81] Hammer, N. e McLeod, D. Data Base Description with SDM: A Semantic Data Model - ACM TODS, vol. 6, no. 3, 1981.
- [MAP87] Moura A.M.C. , Aguiar Filho O.F. e Pereira Neto F.G. - Desenvolvimento de um Banco de Dados Inteligente Utilizando Técnicas de Inteligência Artificial - IV Simpósio Brasileiro de Inteligência Artificial, anais, 1987.
- [MPN87] Moura A.M.C. e Pereira Neto F.G. - Utilização de Técnicas de Inteligência Artificial para Modelização Conceitual de Bco de Dados - SUCESU 87, anais, 1987.
- [MYL78] Mylopoulos, J. et alii. A Preliminary Specification for TAXIS. Technical Report CCA. 1978.
- [SMI77] Smith, J.M. e Smith, D.C.P. Data Bases Abstractions Aggregation and Generalisation - ACM TODS, 1977.
- [TSI82] Tsichritzis, D.C. & Lochovsky, F.H. Data Models. Prentice Hall. 1982.