

SUPORTE DE BANCOS DE DADOS PARA SISTEMAS DE HIPERTEXTOS

Marcos Roberto S. Borges

Núcleo de Computação Eletrônica

e

Instituto de Matemática

Universidade Federal do Rio de Janeiro

Caixa Postal 2324

20001 - Rio de Janeiro

e-mail: borges@ufrj.bitnet

RESUMO

Sistemas de hipertextos e de bancos de dados guardam entre si uma relação muito estreita no que diz respeito ao armazenamento e tratamento de grandes volumes de informação. Apesar disto, muito poucas das soluções adotadas para bancos de dados tem sido adotadas em sistemas de hipertextos, o que parece um contrassenso, já que um hipertexto é um banco de dados. Este artigo estuda esta questão e analisa como as soluções já consolidadas na área de bancos de dados podem ser incorporadas a uma arquitetura de sistemas de hipertextos.

ABSTRACT

Hypertext and database systems have much in common in respect to storage and management of large volumes of data. In spite of that, few are the solutions already adopted in databases that are being applied to hypertext systems. This article addresses this issue and describes how these solutions can be applied to hypertext systems. It also analyses the adoption of an architecture for hypertext systems similar to that proposed by ANSI/SPARC to database management systems.

Palavras-chave: Hipertextos, Hiperbase, Bancos de Dados

I - INTRODUÇÃO

Hipertextos/Hipermídia são termos que hoje já se consolidaram como denominação de uma área própria da ciência da computação. A ideia é considerada antiga [5] mas só recentemente tem recebido maior atenção por parte de pesquisadores. Vários sistemas de diferentes portes e características foram lançados no mercado e muitos mais protótipos experimentais estão em desenvolvimento principalmente em universidades e centros de pesquisa [6, 11, 12, 13].

Esta perspectiva de disseminação generalizada do uso começou a levantar questões sobre os diversos aspectos da arquitetura e implementação de tais sistemas, em especial no suporte ao tratamento de grandes volumes de dados e na interface homem máquina, fatores considerados cruciais no desenvolvimento de sistemas de hipertextos/hipermídia [8, 9, 10, 16].

Há obviamente uma grande diferença entre a implementação de sistemas de uso pessoal, quase experimental, em micro-computadores e aqueles destinados a um uso mais real, ou seja multi-usuário e/ou para máquinas de maior porte como estações de trabalho e até mainframes, em especial no aspecto do tratamento de grandes volumes de informação. Assim, é de se esperar uma grande correlação entre as áreas de hipertextos/hipermídia e bancos de dados [18].

Em princípio, a correspondência das duas áreas pode ser vista como total. Sendo a área de bancos de dados mais consolidada, era de se esperar que muitas das soluções propostas e adotadas para bancos de dados fossem aproveitadas em sistemas de hipertexto [17]. Isto entretanto não acontece. Há alguns sistemas inclusive que abandonaram o uso de um SGBD para suporte ao sistema de hipertexto preferindo-os por sistemas convencionais de arquivos.

Este artigo pretende examinar uma questão que parece ser das mais importantes quando se analisa a arquitetura de sistemas de hipertexto: A correspondência dos problemas de armazenamento e tratamento da informação com sistemas de gerenciamento de bancos de dados. Aspectos mais consolidados como as soluções adotadas para os controles operacionais - concorrência, restrições de integridade, recuperação de falhas e segurança - são avaliados.

A motivação principal deste artigo é exatamente a definição de uma arquitetura para um sistema de hipertexto de suporte ao Projeto HIPERBASE [4] dentro do ambiente PRADES [1, 2, 3] em desenvolvimento na UFRJ.

O restante do artigo se divide em 3 outras partes. Na primeira examinamos brevemente alguns aspectos de hipertextos. Na segunda discutimos a adaptação das soluções consolidadas em bancos de dados para a sistemas de hipertexto e algumas das alternativas em estudo no Projeto HIPERBASE. Na última parte apresentamos as conclusões do trabalho.

II - CARACTERÍSTICAS DE SISTEMAS DE HIPERTEXTOS/HIPERMÍDIA

Hipertextos tem se desenvolvido como uma área separada da informática, embora esteja claro que ela empresta conceitos de diversas outras áreas já consolidadas. Uma boa definição de hipertexto foi dada por Norman Meyrowitz na palestra de abertura da conferência Hypertext '89 [15]: "*Hipertexto é uma encruzilhada onde vários interesses se encontram*".

Sistemas de hipertextos podem ser "vistos" como um grande repositório de informações de diversos tipos: texto, código executável, gráficos, imagens estáticas ou animadas e som; fragmentadas segundo algum critério. Os repositórios podem ser chamados de

hiperdóculos e o particionamento de nós. Uma outra parte é uma estrutura ou um conjunto de estruturas de ligação entre estes nós, denominadas de ligações. Como as ligações entre os nós são do tipo $M \times N$ torna-se necessário definir mais um elemento básico: os botões que representam funções de associação entre trechos de um nó a outro nó. A Figura 1 ilustra um exemplo de um hiperdocumento.

Dentro de um processo de criação de um hiperdocumento há a etapa que é denominada de autoria que consiste das operações: definição e geração do conteúdo dos nós; marcação dos botões de ligação em cada nó; definição dos nós de destino de cada ligação.

Há ainda dois outros conceitos que fazem parte do processo de autoria: a definição de versões e a definição de visões ou contextos [14]. A primeira permite que se possua num mesmo hiperdocumento várias versões de um mesmo nó. A segunda permite diferentes visões do hiperdocumento como um todo, possibilitando, por exemplo, que somente parte dos nós, botões e ligações possam ser percorridos.

A visão ou contexto de um hiperdocumento tem por objetivo apresentar os nós (e suas ligações) que correspondem a um determinado contexto do hiperdocumento. Ao contrário de bancos de dados entretanto, as visões em hipertextos não permitem operar sobre resultados de combinações ou restrições.

Nós e ligações podem possuir ainda um conjunto de descritores que armazenam dados sobre o tipo de nó, o autor, a data de criação, o número da versão, as visões dos quais participam e outros mais.

Um outro modo de operação de um sistema de hipertexto é aquele que permite ao usuário somente navegar pela estrutura sem poderes para altera-la. Normalmente, quando um usuário entra no ambiente é lhe oferecido um primeiro nó (nó de abertura). A partir daí, a navegação se dá por ativação de um dos botões acionados, por exemplo, por um "click" no "mouse". É comum também oferecer alternativas de desvio comuns a todos os nós, através de um menu de botões que desviam para nós de índice ou mapas.

Podemos verificar nesta fase do artigo que é essencial para um sistema de hipertexto um poderoso suporte de armazenamento de dados, seja isto oferecido por um sistema próprio de arquivos ou um sistema gerenciador de bancos de dados. A correspondência entre os problemas e suas possíveis soluções são os temas abordados na próxima seção.

III - SOLUÇÕES DE BANCOS DE DADOS PARA SISTEMAS DE HIPERTEXTO

Apesar de muitos dos problemas e das soluções adotadas para sistemas de hipertexto serem genuínas, há muita interseção com a área de banco de dados, pois ambas se destinam ao tratamento das informações. Nesta seção procuraremos analisar os aspectos do projeto de sistemas de hipertexto que possuem uma solução correspondente em sistemas de bancos de dados e verificar como esta solução se adaptaria a sistemas de hipertexto.

Quando se projeta um sistema cogitando-se apoiar o armazenamento de dados do sistema de hipertexto em um SGBD, é necessário uma análise mais aprofundada das facilidades oferecidas do SGBD, normalmente projetado para sistemas convencionais, onde a informação é naturalmente estruturada. A flexibilidade do enfoque de hipertextos contrasta com esta rigidez e faz com que muitos projetos abandonem já de início esta possibilidade.

A análise de sistemas de hipertextos aqui realizada se concentra em 4 aspectos principais a saber: a arquitetura do sistema, a modelagem do esquema conceitual do hiperdocumento, os mecanismos de busca e a análise dos controles operacionais.

III.1 - Arquitetura do Sistema

Desde 1978 quando foi inicialmente apresentada a proposta da ANSI/SPARC para uma arquitetura de 3 níveis para sistemas de gerenciamento de banco de dados, passou-se a considerar esta arquitetura como ideal porque proporcionava um isolamento de cada uma das 3 camadas: externa, conceitual e interna, das demais. Assim, no nível conceitual, voltado para o administrador do banco de dados, os objetos e seus relacionamentos são tratados isoladamente dos aspectos de máquina e de cada usuário; no nível interno são definidas as estruturas de armazenamento visando a eficiência da recuperação e atualização dos dados; e finalmente no nível externo são definidas as visões externas de cada grupo de usuários.

Tal como estão definidos a maioria dos sistemas de hipertexto, só há um nível de interação, e em alguns casos com dois enfoques: o da autoria e o de leitura/navegação. Os aspectos conceituais e de armazenamento ou não são visíveis ou são tratados conjuntamente com a definição de nós e ligações.

Propõe-se portanto, uma arquitetura de três níveis, semelhante à adotada em sistemas de bancos de dados a fim de isolar os aspectos conceituais, do usuário específico da máquina, alcançando a chamada independência dos dados. Como em sistemas de hipertexto é usual a separação dos leitores dos autores, sugere-se então a definição dessas duas classes no nível externo. Um esquema mostrando a arquitetura proposta é apresentado na Figura 2.

A partir dessa arquitetura, o ponto de partida para a definição de um hiperdocumento passa a ser a definição conceitual da estrutura do mesmo, na qual a abstração dos tipos de nós e ligações é feita. Para isto é necessário um modelo de dados, assunto discutido a seguir.

III.2 - Modelagem de Dados

Conforme mencionado anteriormente, os elementos básicos para estruturação da informação em um sistema de hipertexto são os nós e as ligações. O resultado desta simplicidade é uma grande liberdade por parte do autor na fragmentação da informação, sem qualquer preocupação com um modelo estrutural pré-definido. Por outro lado, embora possamos identificar facilmente a existência de elementos estruturais como nós e ligações, torna-se difícil capturar a semântica destas mesmas ligações ou do conteúdo dos nós. A estrutura conceitual do enfoque de hipertexto é portanto implícita e não há separação do que é estrutura e da instância dessa estrutura.

Em contraste, sistemas de bancos de dados obrigam a estruturação completa da informação forçando que a limitação de dados à base se dê sob regras rígidas, definidas no processo de modelagem de dados. Essa estruturação proporciona uma plataforma para descrever as instâncias na forma de entidades e relacionamentos genéricos e também para a formulação de consultas.

O que propomos é a adoção do mesmo enfoque para sistemas de hipertexto, mas possibilitando e não obrigando a descrição dos tipos de nós e ligações baseada no significado do seu conteúdo.

Algumas propostas nessa direção já foram apresentadas, embora a maioria delas visando uma aplicação específica, e portanto não possibilitando uma estruturação genérica. Um exemplo é o modelo adotado no projeto gIBIS [7]. No hipertexto gIBIS há 3 tipos de nós e 9 tipos de ligações. Esta estrutura facilita o trabalho de autoria e induz a uma melhor visualização do hiperdocumento. A organização estrutural do gIBIS é mostrada na Figura 3.

O projeto HIPERBASE [4] propõe a generalização da ideia de estruturação através da definição conceitual do hipertexto de maneira similar a de um modelo conceitual de banco de dados. Assim a tarefa de autoria é antecedida pela definição da sua estrutura, quando os tipos de nós, ligações e restrições de integridade aplicáveis são definidas. Um exemplo desta definição para um sistema de controle de cartas e propostas é mostrado na Figura 4.

A adoção deste enfoque significa que os conceitos já consolidados na área de modelagem de dados, tais como generalização, agregação e associação, podem ser utilizados em hipertextos.

III.3 - Mecanismos de Busca

Em sistemas de hipertexto, o mecanismo de busca adotado é a navegação. Como o hipertexto não possui uma estrutura conceitual, a navegação é realizada pelas instâncias dos nós, o que significa que o interface de busca em sistemas de hipertexto é a própria informação e não uma linguagem de consulta como estamos acostumados em sistemas convencionais.

Em bancos de dados convencionais, a busca da informação se dá através de uma consulta, que utiliza o esquema como uma estrutura de suporte para formulação das mesmas. Se por um lado o usuário necessita conhecer a estrutura do banco de dados, por outro, o mecanismo de consultas oferece resultados precisos, sem ambigüidade, e com expressividade semântica obtida pelo esquema conceitual do banco de dados.

Conforme sugerido por Halasz [9], o mecanismo de consulta é uma necessidade para grandes hiperdocumentos, a fim de oferecer um meio mais rápido e preciso na localização dos nós de interesse do usuário. Um sistema de hipertexto que utilize um modelo de dados, conforme sugerido na seção anterior, e adote um SGBD para o tratamento de informações pode ter a disposição todo o mecanismo de consulta do sistema, normalmente em uma linguagem padrão SQL, integrando dessa maneira as duas formas de busca: navegação e consulta. Uma extensão interessante do processo de consulta é a de integrar na linguagem meios de tratar as ligações do hiperdocumento, gerando assim uma interface adicional para lidar com as ligações [4].

III.4 - Controles Operacionais

Um outro aspecto refere-se aos controles operacionais necessários. Controle de concorrência, versões, definições e materialização de visões são apenas alguns exemplos das rotinas que devem ser implantadas a partir do zero quando não se usa um SGBD. Este não é absolutamente um trabalho trivial. Os vários anos de pesquisa destas questões na área de bancos de dados possibilitaram o desenvolvimento de sistemas razoavelmente eficientes. O abandono destas soluções pura e simplesmente não é uma solução inteligente.

Por outro lado reconhece-se a rigidez com que estes mecanismos são implantados em sistemas de bancos de dados atualmente em uso. Quase nenhuma flexibilidade (se alguma) é oferecida a quem define o banco de dados, no trato destas questões. As soluções apresentadas por sistemas existentes está muito voltada para o tratamento de dados convencionais o que torna difícil a tarefa de adapta-los à problemática de sistemas de hipertexto.

O aspecto da definição dos contextos oferecidos aos usuários na navegação se assemelha muito à definição de visões em bancos de dados. A diferença está na maneira em que uma visão é definida. Idealmente, a definição de contextos em hipertextos deveria poder contemplar também o modo em que a definição é feita em bancos de dados. Para isso entretanto, é necessário que os nós possuam tipos e descritores associados, como as entidades são definidas em bancos de dados. No projeto Hiperbase contextos são definidas da mesma forma que visões em um SGBD, isto é, através de cortes horizontais e verticais e associações de objetos. O atributo que é tratado diferentemente é o conteúdo do nó que deixa de apontar (e mostrar os botões) para outros nós que não façam parte da visão [4].

O suporte ao trabalho cooperativo é sem dúvida uma das grandes vantagens dos sistemas de hipertextos. Entretanto, para possibilitar esta condição, mecanismos de controle de concorrência se fazem necessários. Neste caso, todavia, há de se buscar soluções adotadas para bancos de dados não convencionais como CAD, que possuem transações longas e portanto necessitam de mecanismos mais flexíveis.

O aspecto de segurança é uma consequência direta do tipo de uso que se fará do hiperdocumento gerado. Até agora os tipos de aplicações propostas induz muito mais ao uso generalizado da informação do que a um acesso restritivo. Entretanto, espera-se que mecanismos de segurança, desde o nível mais trivial baseado em senhas, até aquele auxiliado pela definição de visões ou contextos restritivos, sejam necessários. Também nesta área há vários avanços alcançados na área de bancos de dados que poderão ser aproveitados.

Até agora não tem havido grande preocupação com o problema de recuperação de falhas em hipertextos devido principalmente a inexistência de restrições de integridade mais completas. Com a implementação de tipos de dados e a definição que norteia a existência e o relacionamento entre estes tipos, fica evidente a necessidade de um mecanismo que restaure a integridade do hiperdocumento quando da ocorrência de uma pane qualquer no sistema.

IV - CONCLUSÕES

Neste artigo procuramos demonstrar a necessidade de discussão de aspectos relativos a utilização de sistemas de gerenciamento de bancos de dados, convencionais ou não, no suporte à arquitetura de sistemas de hipertextos. Esta é uma questão que tem grande relevância na definição e utilização de sistemas de hipertextos para um uso mais geral. O assunto se encontra em estudo dentro do projeto HIPERBASE que visa justamente a unificação dos conceitos de bancos de dados e hipertextos dentro do mesmo ambiente.

Na nossa opinião o desenvolvimento de sistemas de hipertextos para uso mais geral não pode dispensar o suporte de um sistema de gerenciamento de bancos de dados, sob o risco de ou aumentar em muito o tempo de desenvolvimento. Por outro lado, é preciso que os SGBDs ofereçam muito mais flexibilidade do que a disponível nos sistemas atuais, a fim de possibilitar um melhor atendimento aos requisitos diferenciados de sistemas de hipertextos. Aliás, esta reivindicação existe também para outras classes de sistemas não convencionais, como CAD, sistemas de tratamento de imagens, etc. Assim, acreditamos que a linha de orientação a objetos ofereça soluções para estes problemas.

A motivação principal do estudo que está sendo realizado dentro do ambiente PRADES é justamente o desenvolvimento de um sistema de hipertexto para dar suporte ao processo de desenvolvimento de software. O ambiente de hardware atualmente disponível para desenvolvimento do software é um VAX 8810 com o sistema operacional VMS. A

justificativa para esta escolha em detrimento de ambientes tipo PC ou estações de trabalho, é que temos como objetivo desenvolver a ferramenta no mesmo ambiente em que o sistema de informação será implantado.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Borges, M.R.S. Uma análise realista da utilização do enfoque da prototipagem rápida no desenvolvimento de sistemas de informação. Anais do XIII Conferência Latino-americana de Informática, Bogotá, Colômbia, Novembro 1987.
- [2] Borges, M.R.S. Rapid prototyping and 4GLs: The change in developing methods of information systems. Anais do IV International Conference: Computers in Institutions of Education, Mexico City, Mexico, Novembro 1988.
- [3] Borges, M.R.S. Hipertextos: A próxima revolução no processo de desenvolvimento de sistemas, Congresso Nacional da SUCEsu '91, São Paulo, Setembro 1991.
- [4] Borges, M.R.S. Combinando bancos de dados e hipertextos: O projeto Hiperbase. Artigo em preparação, 1992.
- [5] Bush, V. As we may think. Atlantic Monthly 176 (1), Julho 1945, pp. 101-108.
- [6] Conklin, J. Hypertext: An introduction and survey. Computer 20 (9), 1987, pp. 17-41.
- [7] Conklin, J & Begeman M. gIBIS: A hypertext tool for exploratory policy discussion. MCC Technical Report STP-082-88, 1988.
- [8] Garg, P.K. e Scacchi, W. A hypertext system to manage software life cycle documents. IEEE Software (7) 3, Maio 1990, pp 90-98.
- [9] Halasz, F. Reflections on Notecards: Seven Issues for the next generation on Hypermedia Systems, Communications of the ACM, (31) 7, Julho 1988, pp 836-852.
- [10] Halasz, F. Seven Issues Revisited. Keynote Talk, Hypertext '91, San Antonio, Texas, Dezembro 1991.
- [11] Anais da Conferência Hypertext '89. Pittsburgh, PA, Novembro 1989.
- [12] Anais da Conferência Hypertext '91. San Antonio, Texas, Dezembro 1991.
- [13] Lima, M.J.D. Hipertextos e suas aplicações. Projeto fim de curso, Departamento de Computação IM/UFRJ, Março 1989.
- [14] Meira, S. et al. Hipertexto: O projeto do sistema H. Anais do II Simpósio Brasileiro de Engenharia de Software, Recife, Outubro 1989, pp. 152-170.
- [15] Meyrowitz, N. Hypertext - Does it reduce cholesterol, too?, Palestra de abertura do Hypertext '89, Novembro 1989, Pittsburgh, USA.
- [16] Reis, A., Borges, M. e Mendes, S. Um modelo híbrido de frames e hipertexto, SBIA'91, Simpósio Brasileiro de Inteligência Artificial, Brasília, DF, 1991.
- [17] Schultz, H. e Streitz, N. Hyperbase: A hypermedia engine based on a relational database management system, European Conference on Hypertext '90, Paris, 1990.
- [18] Tompa, F.Wm e Raymond, D.R. Database methods for hipertext. Course Notes of Hypertext '91, San Antonio, Texas, Dezembro 1991.

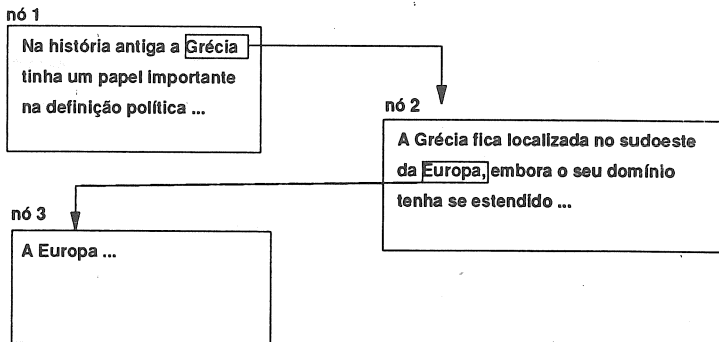


Figura 1 - Exemplo de nós, ligações e botões em um hipertexto

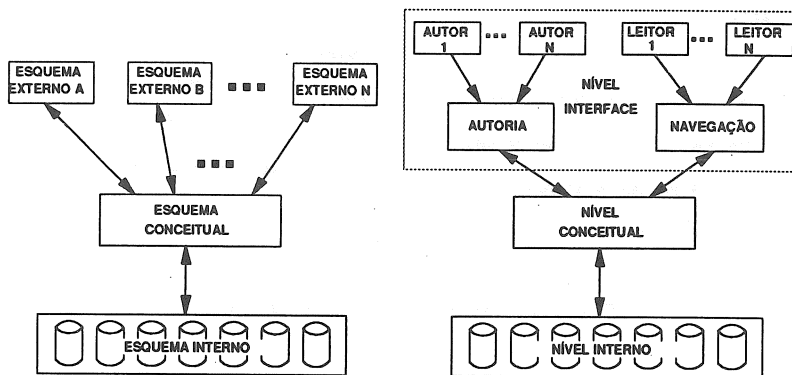


Figura 2 - Arquitetura em 3 níveis de um sistema de hipertexto

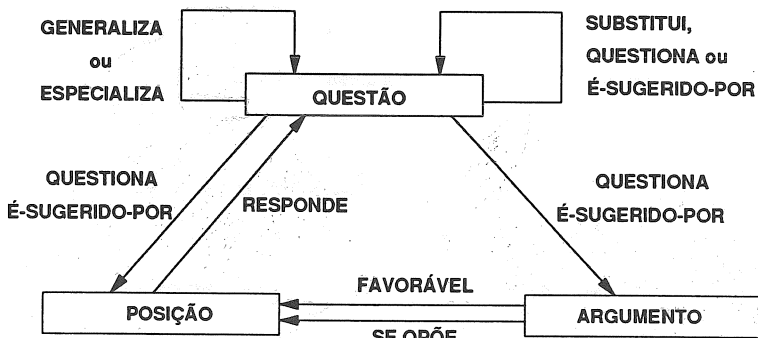


Figura 3 - Tipos de nós e ligações no sistema Gibis

CARTA

CÓD	REM	DEST	DATA	ASSUNTO

PROPOSTA

#PROP	REF	VALOR	DATA

CARTA/PROPOSTA

CÓD	#PROP

Figura 4a - Esquema relacional do banco de dados CARTA/PROPOSTA

CARTA 736/90

EM RESPOSTA A SUA CARTA ENVIADA NO DIA 14/11 ESCLARECEMOS QUE A NOSSA PROPOSTA SEGUE O MESMO PADRÃO DA PROPOSTA 075G/0690 QUANTO AO ITEM TEMPO DE GARANTIA. ESTA INFORMAÇÃO ALIÁS, JÁ TINHA SIDO REITERADA NA NOSSA CORRESPONDENCIA #567/90 DO DIA 22/11/90.

RELACIONAMENTOS OBSERVADOS:

CARTA → RESPONDE → CARTA RECEBIDA
 CARTA → REFERE → OUTRA PROPOSTA
 CARTA → REFERE → CARTA ENVIADA

Figura 4b - Instância da entidade CARTA

CARTA

CÓD	736/90	DATA	27/11/90
REM	ANTONIO DE MORAES		
DEST	FAPERJ - EDUARDO PENNA FRANCA		
ASSUNTO	COMPRA DE REGULADORES DE VOLTAGEM		
CONTEÚDO:			
EM RESPOSTA A SUA CARTA ENVIADA NO DIA 14/11 ESCLARECEMOS QUE A NOSSA PROPOSTA SEGUE O MESMO PADRÃO DA PROPOSTA 075G/0690 QUANTO AO ITEM TEMPO DE GARANTIA. ESTA INFORMAÇÃO ALIÁS, JÁ TINHA SIDO REITERADA NA NOSSA CORRESPONDENCIA #567/90 DO DIA 22/11/90.			

Figura 4c - Entidade CARTA modelada conceitualmente