

RT-MAP - 8106

PROJETO DE BANCOS DE DADOS  
USANDO MODELOS CONCEITUAIS

V.W. Setzer

R. Lapyda

Este relatório Técnico complementa o  
RT-MAP-8103. Ambos substituem o RT-  
MAP-8004 ampliando os conceitos ali  
expostos

JUNHO 1981

## 1. INTRODUÇÃO

Um dos grandes problemas no projeto de bancos de dados usando sistemas comerciais, também chamados "pacotes", é o perigo de se limitar a visão da realidade, reduzindo-a a certas características dos modelos seguidos pelos sistemas em questão. Por exemplo, o uso de pacotes que empregam o modelo hierárquico, como o IMS / DAT/77/ e o SYS-2000 /MRI 72/, pode condicionar os usuários a encararem todas as estruturas de seu problema como hierárquicas. Ou, na melhor das hipóteses, o usuário tem consciência dessa limitação e deve introduzir no projeto de seu banco de dados uma série de artifícios, tais como arquivos auxiliares ou redundâncias, de maneira que o modelo final adapte-se às restrições ou à eficiência do pacote. O projetista pode tender a ficar muito preocupado com esses artifícios, criando um modelo que deixa de ter uma clara correspondência com a realidade. Uma consequência desse fato pode consistir na dificuldade de introduzir alterações no modelo. Para contornar esses inconvenientes, consideramos interessante o uso de um modelo conceitual que não sofre restrições devidas a problemas de implementação. Ele deve ser uma abstração do mundo real em um nível puramente formal, porém com características adequadas a sistemas de informação. Nesse sentido, escolhemos o Modelo Entidade-Relacionamento, "Entity-Relationship Model" (MER), de Chen /CHE 76/ como modelo conceitual. Não entraremos em detalhes desse modelo, fazendo apenas uma breve introdução ao Diagrama Entidade Relacionamento (DER), que constitui o cerne do MER. Serão abordadas apenas as principais características do DER. A seguir, fazemos breve introdução ao modelo de dados do sistema ADABAS /SOF 71/, inclusive com algumas considerações sobre as suas linguagens de consulta, e apresentamos um diagrama representativo desse modelo. Finalmente, apresentamos uma metodologia para conversão do DER para os arquivos ADABAS.

Na medida das possibilidades e do espaço, empregaremos uma descrição informal.

## 2. O MODELO ENTIDADE-RELACIONAMENTO

O modelo introduzido por Chen considera que o mundo real pode ser representado conceitualmente como uma coleção de conjuntos de entidades (CE) e de conjuntos de relacionamentos (CR) existentes entre entidades. Não se pode definir o que vem a ser uma entidade; a grosso modo, trata-se da abstração de qualquer objeto, ser, ou evento da realidade. Os relacionamentos são associações formais entre elementos de CEs. Por exemplo, podemos ter os CEs ALUNOS e DISCIPLINAS, e um CR HISTÓRICO ESCOLAR. Cada elemento deste associa um elemento do conjunto ALUNOS a um elemento do conjunto DISCIPLINAS. O DER correspondente a esses tres conjuntos seria o da fig. 1.



fig.1

Nesse diagrama há uma notação adicional, que é lida  $m:n$  ("m para n"), e que caracteriza o tipo do CR HISTÓRICO ESCOLAR. Nesse caso, fica explícito que um elemento de ALUNOS, isto é, um aluno, pode ser relacionado ou associado a mais do que um elemento de DISCIPLINAS e vice-versa. Isto é, interpretando o CR como representando a informação de quais alunos concluíram quais disciplinas, podemos deduzir do tipo  $m:n$  que cada aluno pode ter concluído um número qualquer (não limitado) de disciplinas, e que cada disciplina pode ter sido concluída por um número qualquer de alunos.

Vamos estender a notação de Chen, introduzindo números naturais quaisquer em lugar de  $m$  e/ou  $n$ , e que pode melhor ser explicada por meio do exemplo da fig.2.

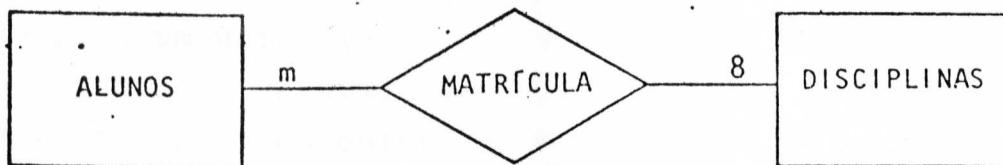


fig.2

Podemos associar ao DER da fig.2 a interpretação de que cada aluno pode estar matriculado em até 8 disciplinas, cada disciplina pode ter um número qualquer de alunos nela matriculado.

Em particular, pode-se ter CRs dos tipos 1:n e 1:1. Note-se que os CRs MATRÍCULA e HISTÓRICO ESCOLAR podem coexistir, isto é, dois CEs podem ser relacionados por mais de um CR.

A cada CE ou CR pode-se associar funções (no sentido matemático) que levam cada elemento desses conjuntos a um só elemento de um conjunto de valores. Essas funções são chamadas de atributos; sua representação diagramática é dada na fig.3, conforme Chen /CHE 77/.

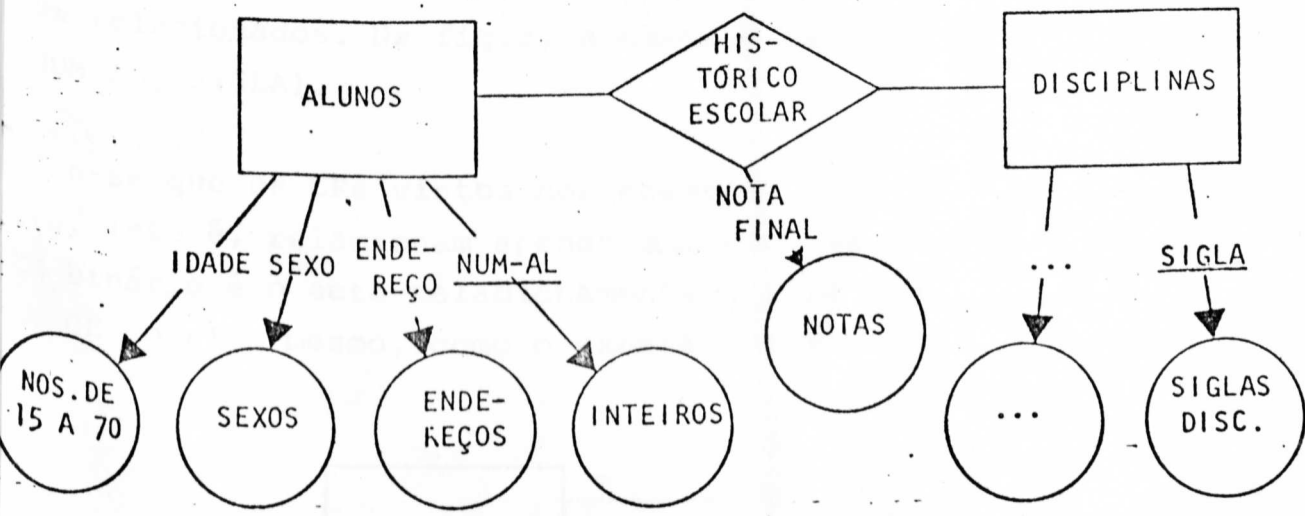


fig.3

Na fig.3, são atributos IDADE, SEXO, ENDEREÇO, NUM-AL, NOTA FINAL; são conjuntos de valores NOS.DE 15 a 70, SEXOS, ENDEREÇOS, INTEIROS, NOTAS, SIGLAS DISC. Sua interpretação pode ser entendi

da como representando propriedades de cada entidade ou relacionamento. Assim, IDADE é uma função que leva um elemento do conjunto de alunos a um único valor que é um número natural, atribuindo dessa maneira um determinado valor de idade para cada aluno. Da mesma forma, podemos entender NOTA FINAL como uma nota associada a cada par aluno-disciplina, representando com qual nota o aluno concluiu uma determinada disciplina. Note-se que esse atributo origina-se no CR, e não é característica (atributo) de elementos de ALUNOS isoladamente, e nem de elementos de DISCIPLINAS.

Na fig.3 os atributos NUM-AL e SIGLA são especiais, pois constituem-se em *chaves primárias*, ou simplesmente *chaves* dos CEs a que são aplicados. Uma chave é um atributo cuja função é uni-única (isto é, 1 para 1). Assim, para cada elemento do CE a chave leva a um único elemento do conjunto de valores e a esse elemento só corresponde um elemento do CE segundo essa chave. Por exemplo, cada aluno tem apenas um número de aluno e vice-versa. Os valores das chaves servem, assim, para identificar os elementos do CE correspondente. Pode acontecer que uma chave tenha que ser formada por mais do que um atributo. As chaves serão indicadas sublinhando-se os nomes dos atributos.

Os CRs também tem chaves; estas são constituídas das chaves dos CEs relacionados. Na fig.3, a chave de HISTÓRICO ESCOLAR é o par (NUM-AL, SIGLA).

Note-se que os CRs vistos nos exemplos acima são de classe *binária*, isto é, relacionam apenas dois CEs. Um caso particular de CR binário é o auto-relacionamento binário, um CR que relaciona um CE consigo mesmo, como o exemplo da fig.4.

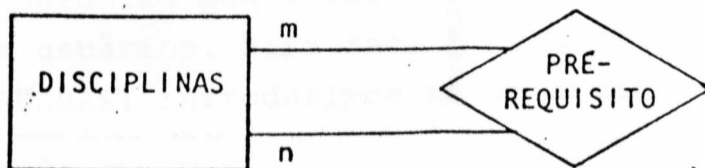


fig.4

Pode-se ter classes de CR relacionando mais do que dois CEs como por exemplo a classe *ternária*.

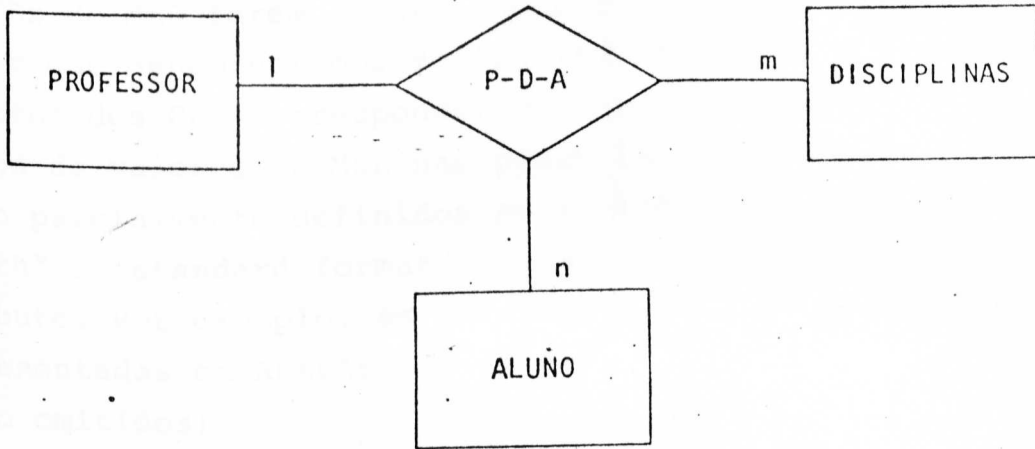


fig.5

Na fig.5 mostramos um CR ternário do tipo 1:m:n. Interpretaremos esse CR como sendo o fato de que a cada par aluno-disciplina corresponde no máximo um professor, isto é, cada aluno só pode ter um professor por disciplina. Por outro lado, um aluno pode estar cursando com um certo professor um número qualquer de disciplinas, e um professor pode ministrar uma certa disciplina a qualquer número de alunos. Os CRs ternários podem ainda ser do tipo 1:1:n, m:n:p etc.

Existem interessantes extensões do MER (/S-S 79/, /S-N 79/) que não serão utilizadas neste trabalho, pois é nossa intenção permanecer no mais simples nível, simplificando tanto a compreensão do método a ser exposto, quanto a sua execução em casos práticos.

### 3. O MODELO DE DADOS DO SISTEMA ADABAS

Nesse item, faremos uma introdução ao sistema ADABAS, principalmente no que concerne ao seu modelo lógico de dados, isto é, como visto pelos usuários. Para facilitar a conversão do MER para o modelo do ADABAS, introduzimos novas noções neste último.

### 3.1. E-ARQUIVOS

Aos CEs do MER faremos corresponder "files" ou arquivos, do ADABAS, e que denominaremos de "entity files" ou E-arquivos. Aos atributos dos CEs corresponderão "fields" ou E-atributos. Os conjuntos de valores do MER não podem ser totalmente especificados, sendo parcialmente definidos pelas características "standard length" e "standard format", e que denominaremos de *formato* do atributo. Por exemplo, as características do CE ALUNOS podem ser representadas em ADABAS por meio da declaração (alguns detalhes serão omitidos):

#### ALUNOS

01, NA, 6, F, DE	(número do aluno)
01, ID, 2, F	(idade)
01, EN, 25, A	(endereço)
01, SX, 1, A	(sexo)

Cada linha depois do nome ALUNOS do E-arquivo corresponde a um E-atributo. Nestes, temos especificados, pela ordem: um nível do atributo, como em COBOL ou PL/I; o identificador do atributo; o formato, composto das especificações de comprimento e tipo (F corresponde a inteiro, A a alfanumérico); o primeiro atributo contém uma especificação 'DE', que o caracteriza como uma *descriptor*, passando com isso a ser um atributo que pode ser usado nas linguagens de consulta dentro dos *critérios de seleção* de registros.

Note-se que nem sempre há uma correspondência total entre os conjuntos de valores do MER e os formatos dos atributos do ADABA. De fato, no caso do atributo ID, não é possível especificar os números naturais de 15 a 70 como constituindo o conjunto de valores desejado; da mesma maneira não é possível restringir o formato de SX somente aos valores 'M' e 'F'.

Em alguns CEs e em todos os CRs do MER, ocorrem chaves compostas de mais de um atributo. Existe uma correspondência direta destas chaves compostas em ADABAS: são os denominados "super-descriptors", que envolvem mais do que um atributo de um arquivo

e funcionam, do ponto de vista dos critérios de seleção, como um descritor. No restante deste trabalho, consideraremos descritores em geral, sem especificar se são "superdescriptors" ou não. A representação diagramática de um E-arquivo é apresentada na fig.6.

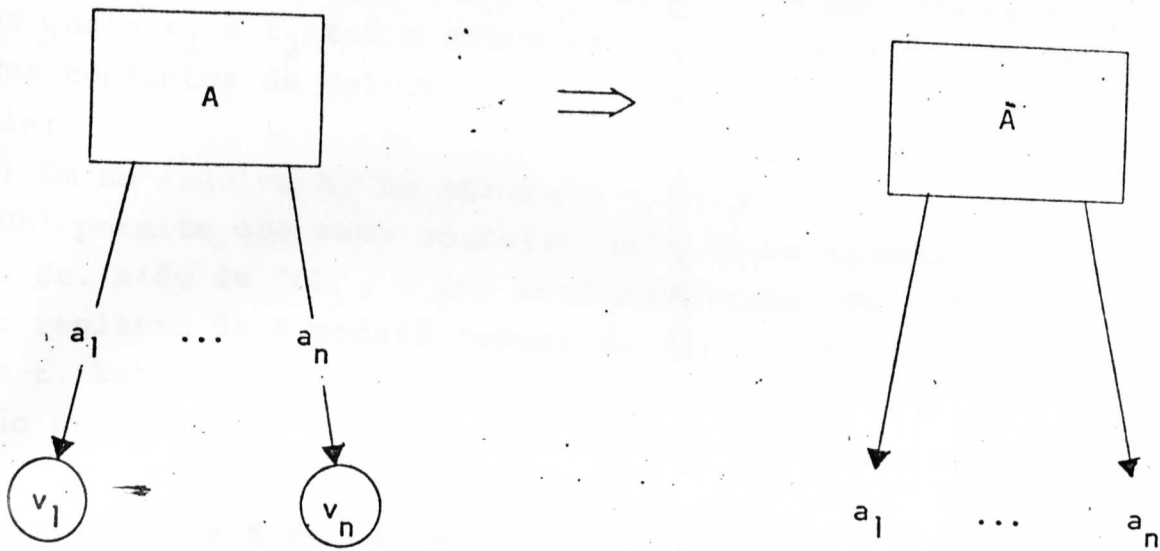


fig.6

### 3.2. CONEXÕES

Os CRs do MER serão representados por construções do ADABAS que denominaremos de *conexões*, que podem ser de duas classes: *implícitas* ou *explícitas*.

#### 3.2.1. CONEXÕES IMPLÍCITAS

Essas conexões são implementação de relacionamentos do MER que usam apenas "couplings" ou *acoplamentos*, entre E-arquivos do ADABAS. Esses acoplamentos tem as seguintes características:

- a) Acoplamentos podem ser feitos entre dois e somente dois arquivos distintos, isto é, são sempre binários;
- b) Dois arquivos só podem ser acoplados por meio de um único acoplamento. Cada arquivo pode ser acoplado a no máximo 80 outros arquivos;
- c) Se os arquivos  $A_1$  e  $A_2$  estão acoplados, existe uma associação implícita entre registros de  $A_1$  e registros de  $A_2$ ;
- d) O acoplamento entre dois arquivos  $A_1$  e  $A_2$  é feito chamando-se

dinamicamente uma rotina utilitária do sistema, à qual são fornecidos os identificadores de  $A_1$  e  $A_2$ , bem como os identificadores de um atributo de cada um, cuja declaração tenha a especificação 'DE';

e) Se um acoplamento entre dois arquivos  $A_1$  e  $A_2$  é feito através dos atributos  $t_1$  de  $A_1$  e  $t_2$  de  $A_2$ , uma associação implícita é estabelecida entre cada registro de  $A_1$  e os registros de  $A_2$  para os quais  $t_1$  e  $t_2$  tem o mesmo valor, e vice-versa. A intersecção dos conjuntos de valores de  $t_1$  e  $t_2$  não deve, portanto, ser vazia;

f) Em um arquivo A, um atributo t declarado simplesmente como 'DE' permite que cada registro de A tenha apenas um valor para t. Se, além de 'DE', t for declarado como 'MU' ("multiple"), cada registro de A poderá conter de zero a 191 valores diferentes de t. Esta característica do ADABAS será empregada na implementação simplificada de certos CRs do MER, como veremos no item 4.

Nas figs. 7 e 8 damos exemplos de conexões implícitas, sem atributos especificados como sendo do tipo 'MU' e, na fig.9, com essa especificação

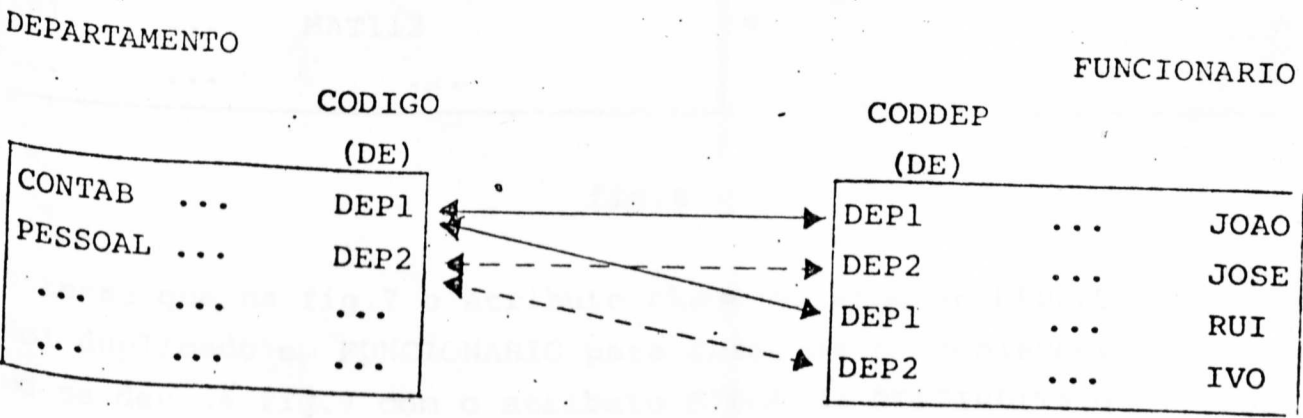


fig. 7

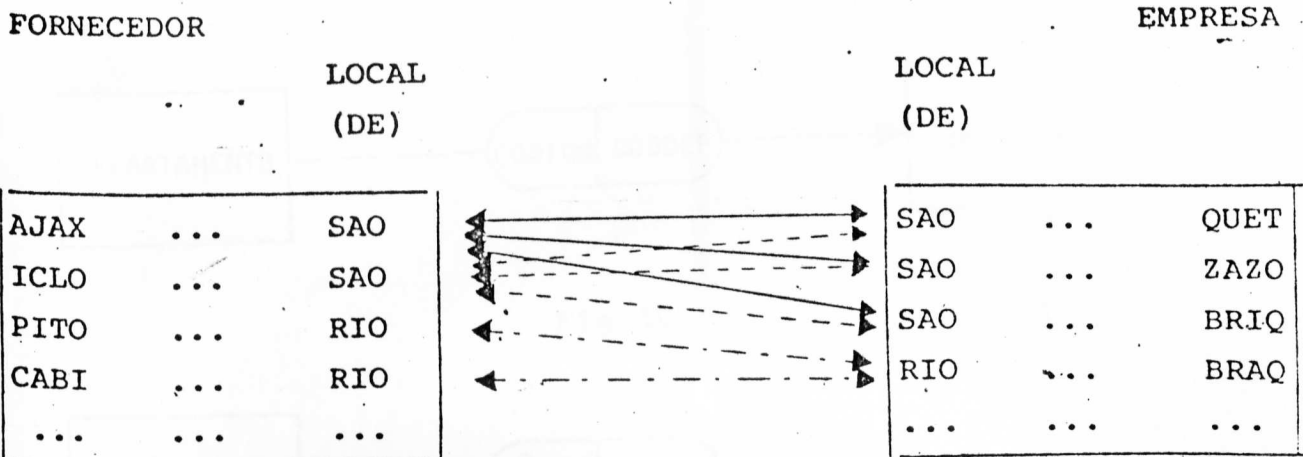


fig.8

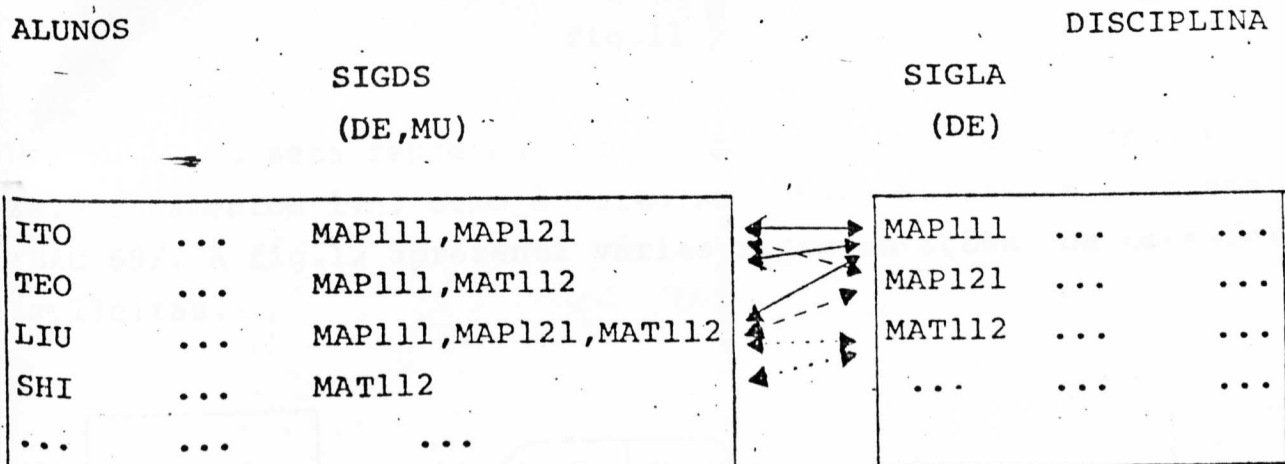


fig.9

Note-se que na fig.7 o atributo chave do arquivo DEPARTAMENTO foi duplicado em FUNCIONARIO para obter-se o acoplamento; o mesmo se deu na fig.9 com o atributo SIGLA de DISCIPLINA que foi duplicado como o atributo 'MU' SIGDIS de ALUNOS. Já na fig.8, os atributos de nome LOCAL são próprios dos arquivos FORNECEDOR e EMPRESA, não tendo havido necessidade de duplicação. As conexões implícitas estão sujeitas às restrições dos acoplamentos, como por exemplo o fato de serem sempre binários. Introduzimos, nesse ponto, uma representação gráfica para os acoplamentos. As figs. 10 e 11 mostram essa representação para os exemplos das figs.7 e 9, respectivamente. Note-se que no bloco arredondado aparecem os descritores usados no acoplamento. A especificação 'MU' é representada por um '\*'.

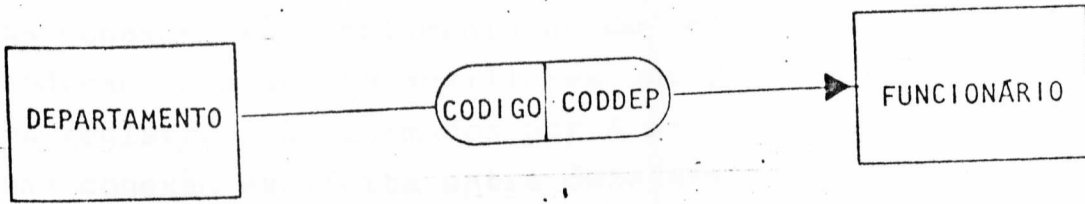


Fig.10

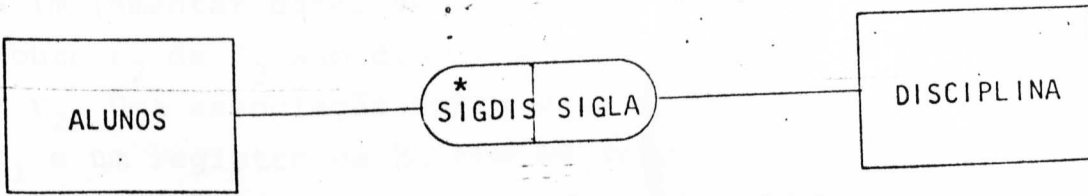


fig.11

Na fig.10, a seta representa uma implementação de um conjunto de relacionamentos 1:n, como é feito nos "Data Structure Diagrams" /BAC 69/. A fig.12 apresenta várias representações de conexões implícitas.

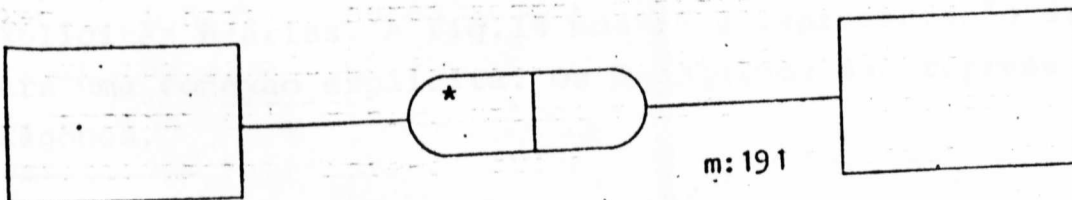
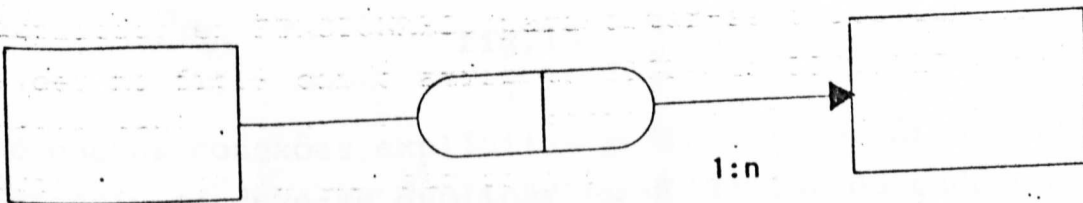
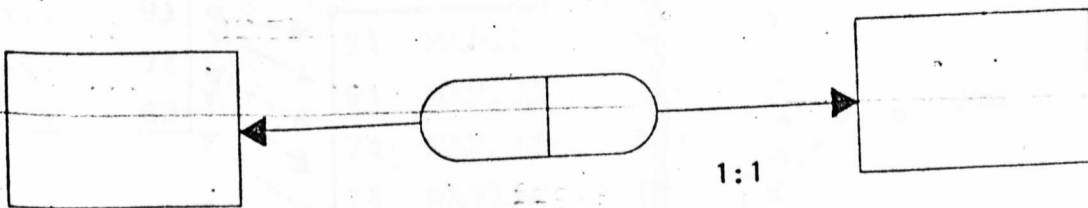


fig.12

### 3.2.2. CONEXÕES EXPLÍCITAS

Essas conexões são implementação de relacionamentos do MER pela introdução de arquivos auxiliares que denotaremos por *R-arquivos* cujos registros são formados por *R-atributos*. Para estabelecer-se uma conexão explícita entre dois arquivos  $E_1$  e  $E_2$  (que em geral serão E-arquivos, com uma exceção), é introduzido um R-arquivo R e dois acoplamentos, um entre R e  $E_1$  e o outro entre R e  $E_2$ . Para implementar esses acoplamentos, um atributo  $t_1$  de  $E_1$  e um atributo  $t_2$  de  $E_2$  são duplicados em R, constituindo os atributos  $r_1$  e  $r_2$ . Uma associação entre um registro de  $E_1$  com um valor  $v_1$  de  $t_1$  e um registro de  $E_2$  com um valor  $v_2$  de  $t_2$  é estabelecida pela introdução de um registro de R com valores de  $r_1$  e  $r_2$  iguais a  $v_1$  e  $v_2$  respectivamente. Na fig.13 damos um exemplo dessa conexão, onde  $E_1$  corresponde a ALUNOS,  $E_2$  a DISCIPLINAS, R a HISTÓRICO ESCOLAR,  $t_1$  a NUM,  $t_2$  a SIGLA,  $r_1$  a NUAL e  $r_2$  a SIGD.

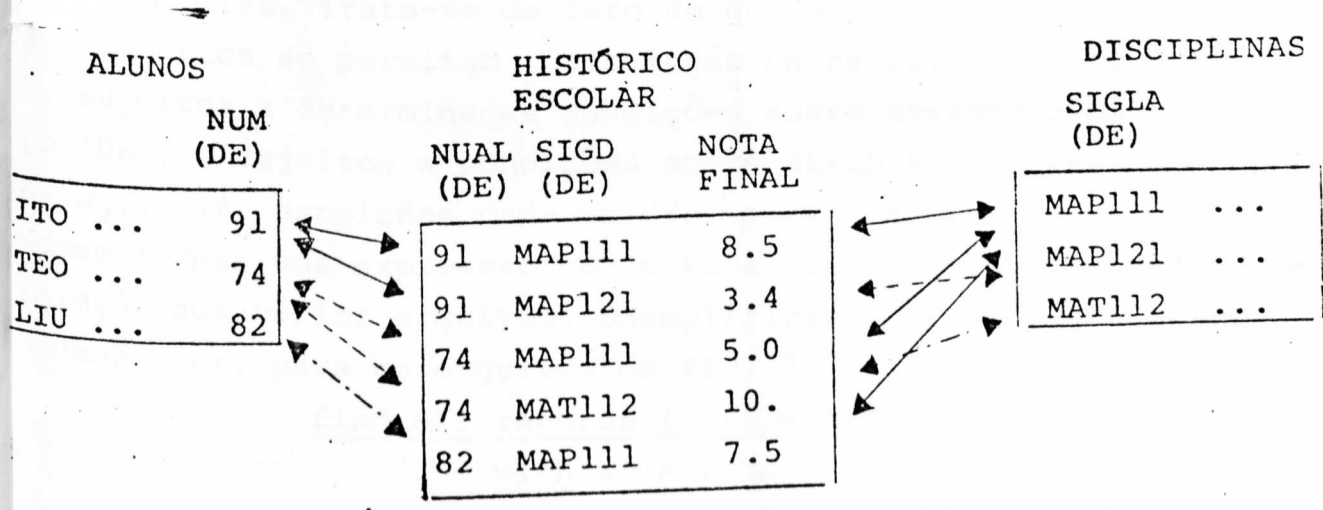


fig.13

Note-se que as conexões explícitas podem ser ternárias, ligando três E-arquivos; deve-se duplicar um descritor de cada um destes no R-arquivo da conexão, e introduzir-se três acoplamentos do R-arquivo para cada um dos E-arquivos. Em geral, podemos ter conexões explícitas n-árias. A fig.14 mostre a representação diagramática para uma conexão explícita. Os R-arquivos são representados por hexágonos.



fig.14

### 3.3 LINGUAGENS DE CONSULTAS

Não entraremos aqui nos detalhes das várias linguagens de consulta do ADABAS. Importa-nos descrever uma característica fundamental da linguagem básica de consulta que usa linguagens hospedeiras "host languages"), bem como da linguagem de consulta direta (query language") ADASCRIPIT e da linguagem independente NATURAL devido ao impacto na simplicidade de formulação e na eficiência das consultas. Trata-se do fato de que as consultas que utilizam acoplamentos só permitem a seleção de registros de um arquivo A, sujeitos a determinadas condições sobre atributos de A do tipo 'DE', e sujeitos a condições sobre atributos de arquivos  $A_1, A_2, \dots, A_n$  acoplados cada um *diretamente* a A; essas condições devem formar uma expressão conjuntiva (isto é, do tipo "e") em relação aos vários arquivos. Exemplificando, usando ADASCRIPIT, podemos ter, para os arquivos da fig.13:

```
find all records in file HISTORICO-ESCOLAR  
with NOTA-FINAL < 5.0 and  
coupled to file ALUNOS with IDADE > 30  
and SEXO = 'M'  
and coupled to file DISCIPLINAS  
with PERIODO = 'NOTURNO'
```

Note-se que supomos a existência dos atributos IDADE e SEXO em ALUNOS e PERIODO em DISCIPLINA.

Antes da palavra reservada coupled só pode ser usada a conjunção and. Assim, tanto não se pode usar outros conectivos lógicos entre acoplamentos, como não se pode *encadear* acoplamentos em uma só consulta, como por exemplo "ache registros de A acoplados a

registros de B, que por sua vez estão acoplados a registros de C". Mais concretamente, não seria possível formular, para os arquivos da fig.13, a consulta "obter os registros de alunos que tiveram nota final menor do que 5.0 em disciplinas do período noturno" (novamente consideramos que DISCIPLINAS contém um atributo PERÍODO).

#### 4. CONVERSÃO DO MER PARA O ADABAS

Neste ítem, introduziremos um procedimento para conversão sistemática do DER para o ADABAS. Devemos, para isso, adotar um critério geral, que será o seguinte: os esquemas de conversão serão tais, que permitam buscas na linguagem básica (cuja restrição se reflete nas outras linguagens) usando, sempre que possível, somente acoplamentos, isto é, sem que seja necessário colecionar valores de algum atributo de um arquivo para posteriormente usá-los como valores de descritores na busca de registros do mesmo ou de outro arquivo. Em termos do ADABAS isto equivale a trabalhar, quando possível, apenas com as ISNs que são, essencialmente, endereços dos registros. Esse procedimento consta, basicamente, de dois passos: geração e redução.

##### 4.1. GERAÇÃO

4.1.1. Para cada conjunto de entidades T do DER, criar um E-arquivo E, com os campos correspondentes aos atributos de T; a chave de T deve corresponder um campo especificado com 'DE' (fig. 15).

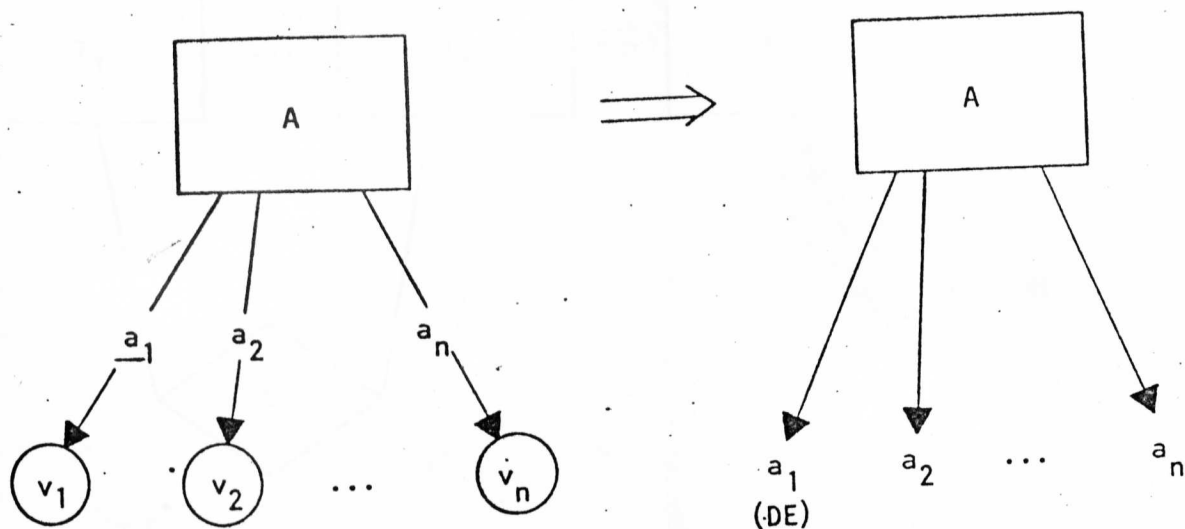


fig.15

#### 4.1.2. Com relação aos conjuntos de relacionamentos:

4.1.2.1. Para cada conjunto de relacionamentos L, que não seja de auto-relacionamentos, criar um R-arquivo R, estabelecendo uma conexão explícita entre os E-arquivos que representam os conjuntos de entidades relacionados através de L. Cada um desses E-arquivos terá sua chave duplicada no R-arquivo criado e os acoplamentos entre os E-arquivos e R serão estabelecidos justamente através dessas chaves e de suas respectivas duplicações.

A fig. 16 mostra a representação diagramática desse tipo de conversão, incluindo o caso geral e os casos de relacionamentos binários 1:1 e 1:n.

4.1.2.2. Para cada conjunto de relacionamentos L, binário e de auto-relacionamentos, que relacione um conjunto de entidades T com ele mesmo:

- criar um arquivo auxiliar X, cujo tipo será denominado X-arquivo. Este arquivo terá apenas um atributo, que será uma duplicação da chave do E-arquivo E, correspondente ao conjunto de entidades T

- criar um R-arquivo R, estabelecendo uma conexão explícita entre E e X; os atributos de L serão introduzidos em R (fig.17).

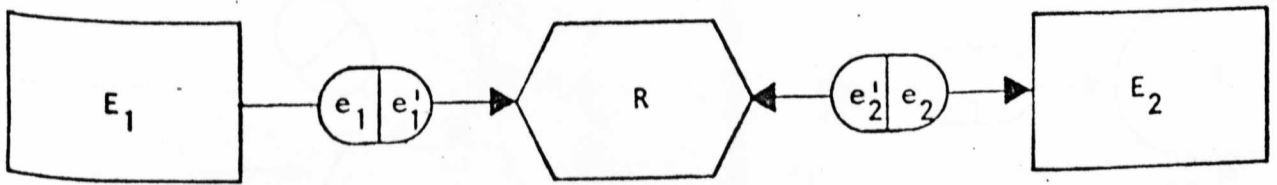
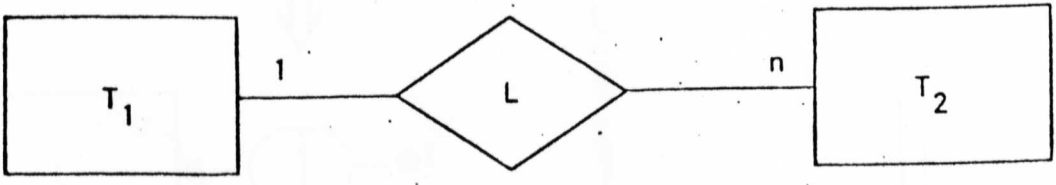
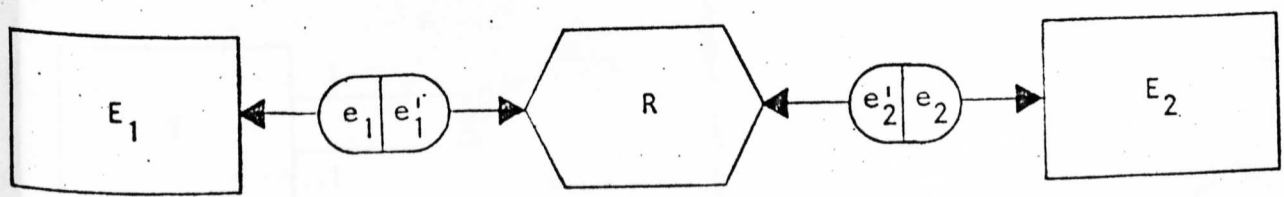
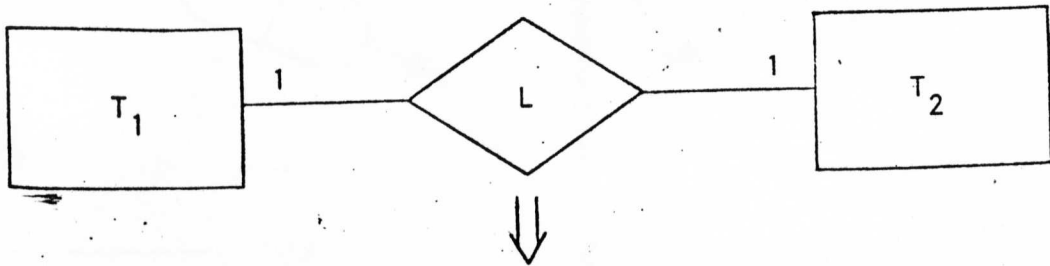
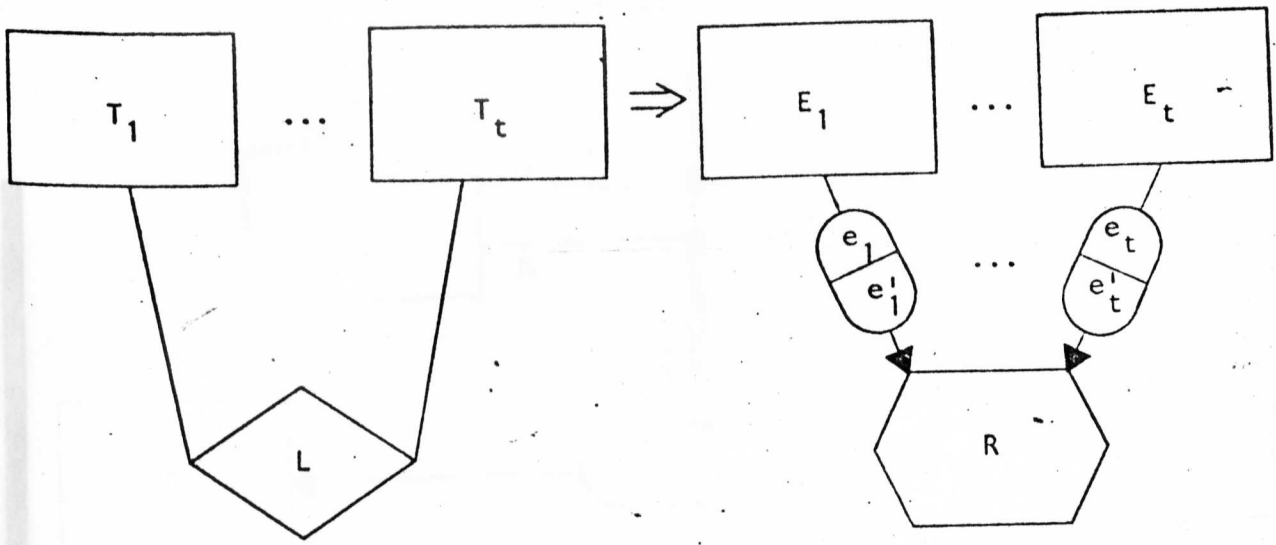


fig.16

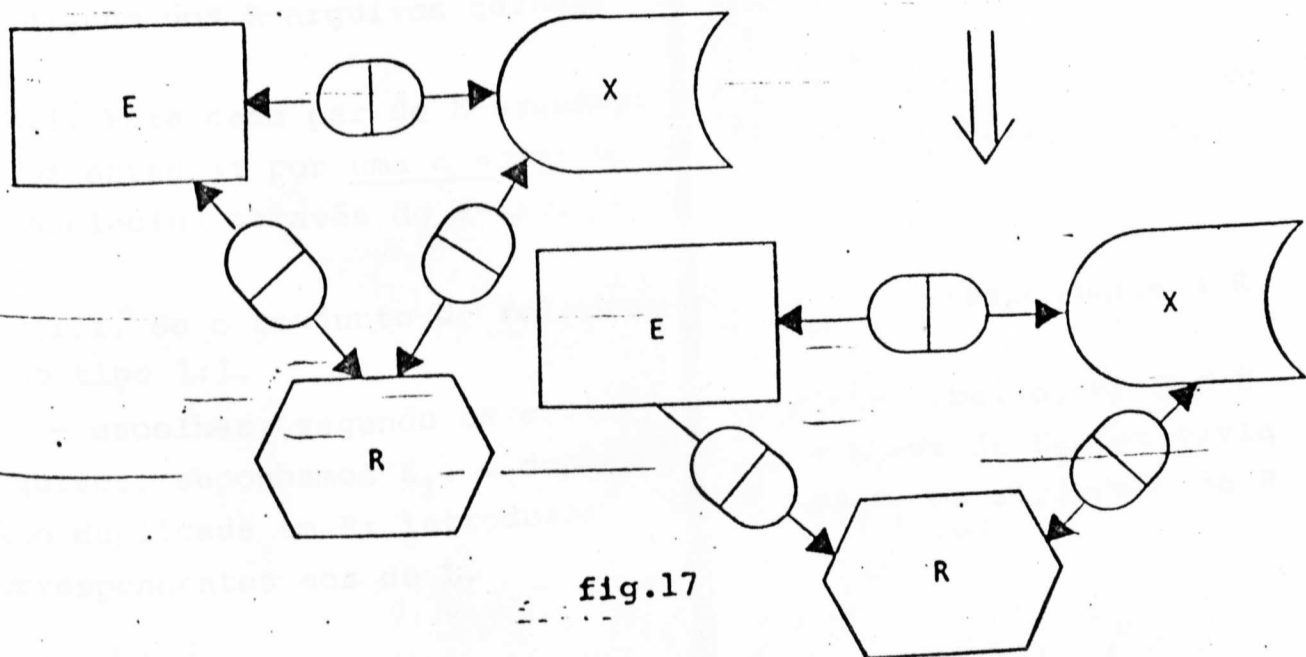
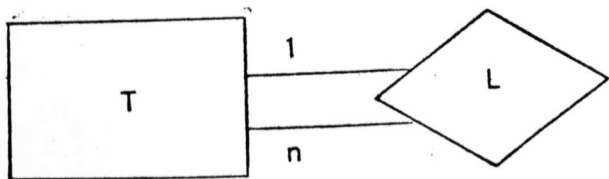
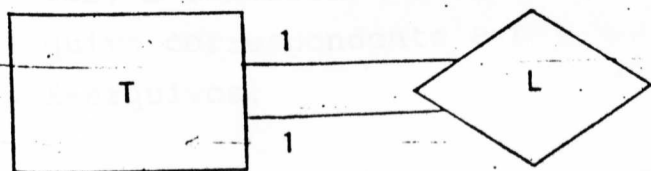
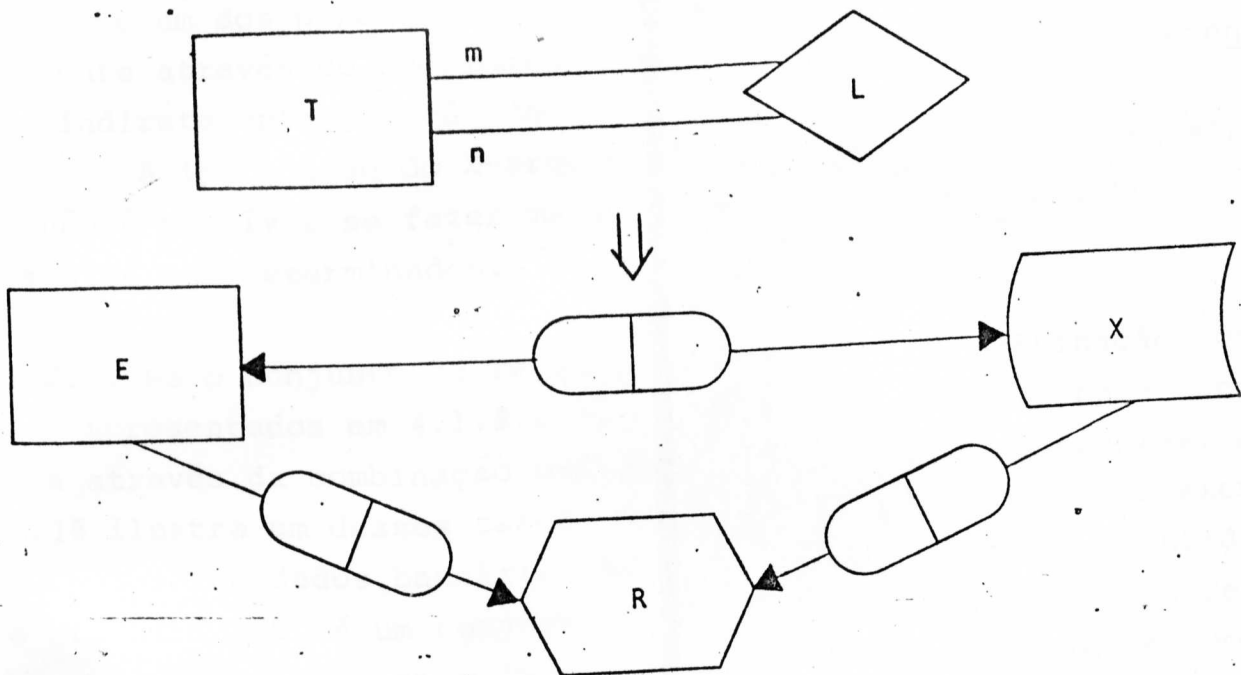


fig.17

Lembrando que no MER diz-se que cada conjunto de entidades tem um *papel* nos relacionamentos em que tomam parte, pode-se dizer que um dos papéis que T representa em L é implementado diretamente através do acoplamento com R; o outro papel é implementado indiretamente através de X.

A introdução do X-arquivo se deve ao fato de que no ADABAS não é possível se fazer mais do que um acoplamento entre dois arquivos determinados.

4.1.2.3. Se o conjunto de relacionamentos L é uma combinação dos casos apresentados em 4.1.2.1. e 4.1.2.2., a conversão pode ser feita através da combinação das duas técnicas correspondentes. A fig.18 ilustra um desses casos, que pode ser associado a um exemplo de base de dados bancária, em que  $T_1$  é um conjunto de entidades CLIENTES e  $T_2$  é um conjunto de entidades EMPRÉSTIMOS; os relacionamentos expressam o fato de que um cliente garante o empréstimo de outro.

De um modo geral, quando se tem auto-relacionamentos n-ários, a conversão é feita através de um acoplamento direto ao R-arquivo correspondente e n-1 acoplamentos indiretos através de n-1 X-arquivos.

## 4.2. REDUÇÃO

A redução consiste na substituição de algumas conexões explícitas por conexões implícitas, com a consequente eliminação de alguns dos R-arquivos gerados no passo anterior.

4.2.1. Para cada par de E-arquivos,  $E_1$  e  $E_2$  que estejam relacionados entre si por uma e somente uma conexão explícita binária, estabelecida através do R-arquivo R:

4.2.1.1. Se o conjunto de relacionamentos L, correspondente a R, é do tipo 1:1.

- escolher, segundo os critérios descritos abaixo, um dos E-arquivos, suponhamos  $E_1$ , e duplicar nele a chave de  $E_2$  que havia sido duplicada em R; introduzir, também, em  $E_1$  os atributos de R correspondentes aos de L.

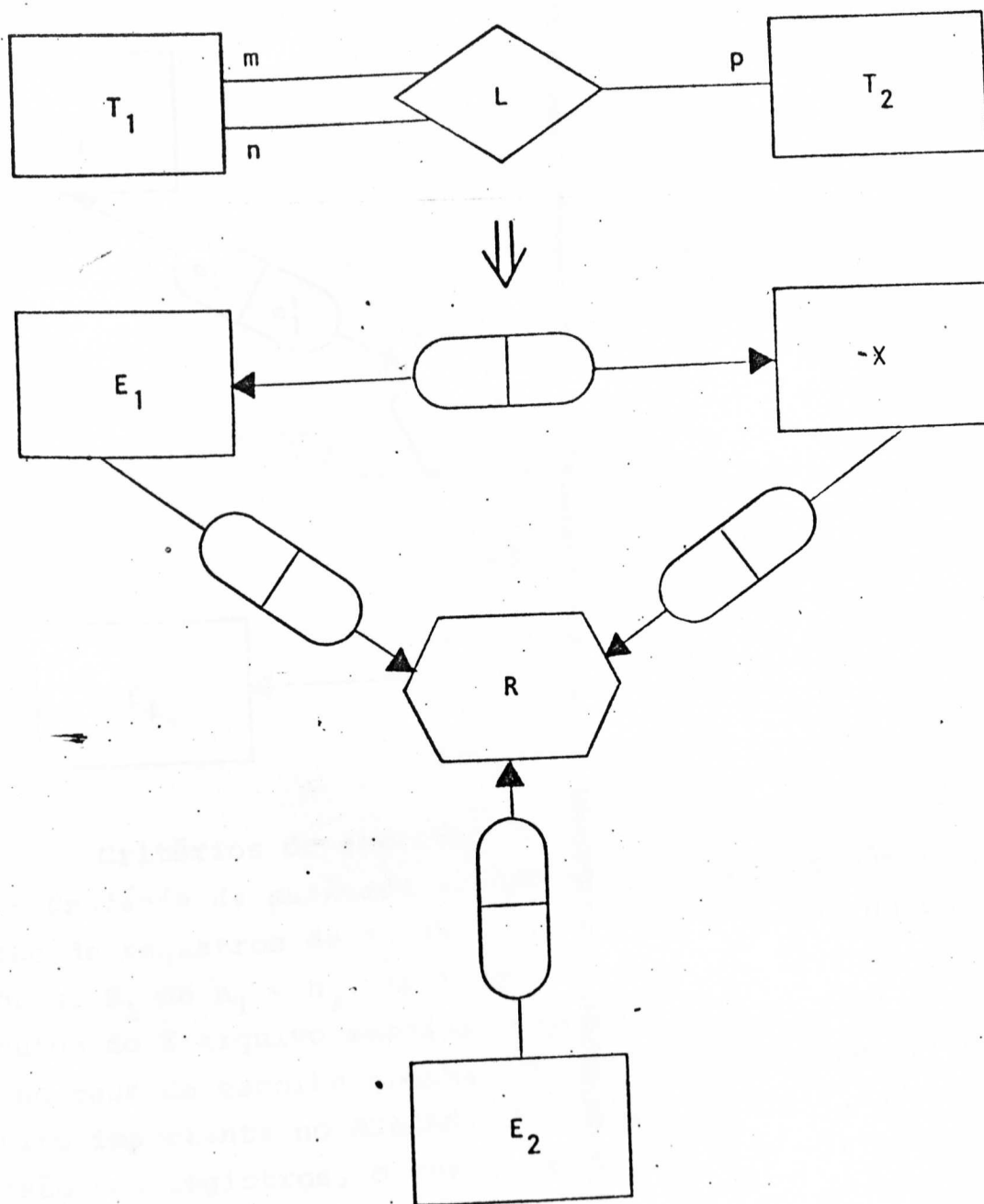
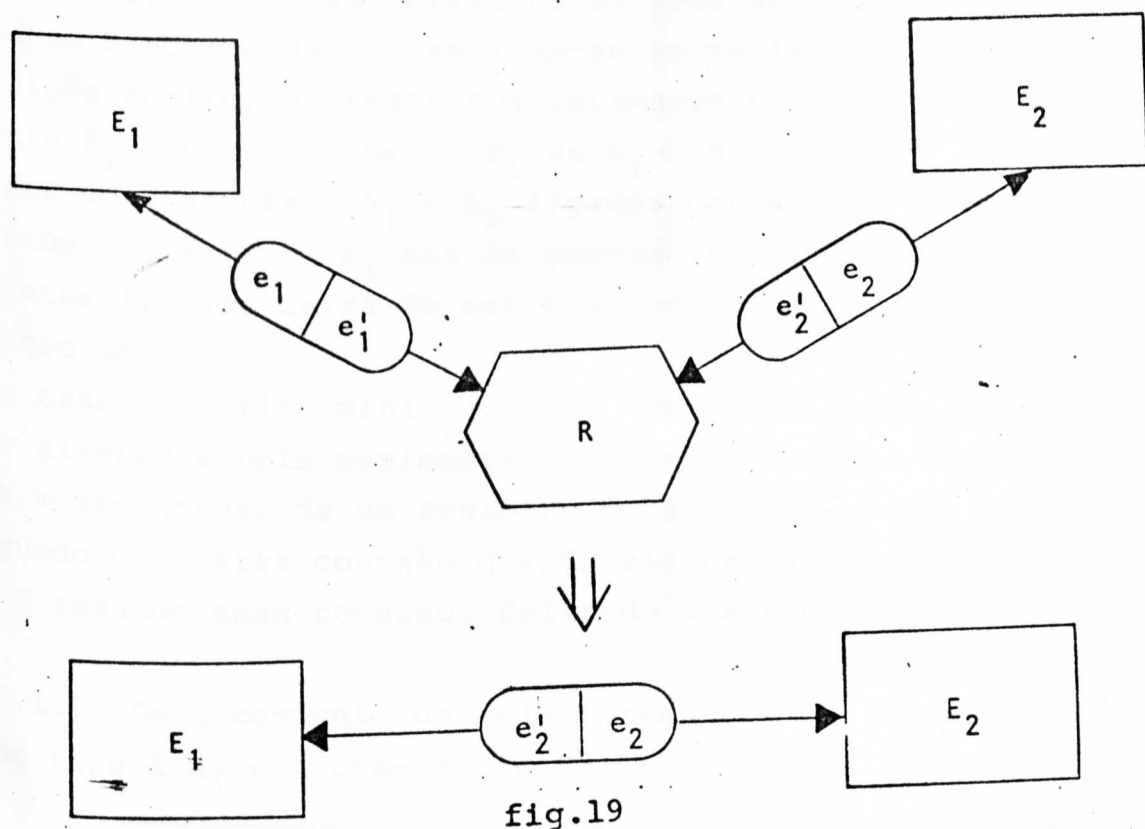


fig.18

- eliminação: o arquivo  $R$ , bem como os acoplamentos associados, são retirados do modelo e, no seu lugar, é estabelecida uma conexão implícita entre  $E_1$  e  $E_2$ , através da chave de  $E_2$  e de sua duplicação em  $E_1$



**Critérios de escolha:**

- *Critério de ausência no CR:* seja  $n_1$  ( $n_2$ ) o número médio estimado de registros de  $E_1$  ( $E_2$ ) que não participam da conexão. Escolhe-se  $E_1$  se  $n_1 < n_2$ . Aplicando-se esse critério, os novos atributos do E-arquivo escolhido terão menos valores vazios do que no caso da escolha oposta. Em particular, esse critério não é muito importante no ADABAS, devido às possibilidades de compressão dos registros, o que elimina destes os atributos com valores vazios.

- *Critério de frequência de consultas:* Para  $n$  consultas ao banco de dados, estima-se em média  $n_1$  ( $n_2$ ) consultas que localizam registros de  $E_1$  ( $E_2$ ) usando-se, entre outros atributos, condições sobre a chave de  $E_2$  ( $E_1$ ) e/ou atributos do CR. Escolhe-se  $E_1$  se  $n_1 > n_2$ . Com esse critério maximiza-se o número de buscas que não usam acoplamento.

- *Critério de alteração da chave:* Para  $n$  consultas ao banco de dados estimam-se em média  $n_1$  ( $n_2$ ) atualizações das chaves de  $E_1$  ( $E_2$ ). Escolhe-se  $E_1$  se  $n_1 > n_2$ . Com esse critério, minimiza-se o número de consultas que alteram ambos os arquivos simultaneamente. Com isso, não só aumenta-se a

eficiência, mas também evitam-se problemas de integridade.

- *Critério de estabilidade de presença no CR:* Para  $n$  consultas ao banco de dados, estimam-se em média  $n_1$  ( $n_2$ , saídas (cf. de definição abaixo) de registros pertencentes a  $E_1$  ( $E_2$ ) da conexão entre  $E_1$  e  $E_2$ . Escolhe-se  $E_1$  se  $n_1 < n_2$ .

Dados dois arquivos  $A_1$  e  $A_2$  ligados por alguma conexão, dizemos que um registro de  $A_1$  sai da conexão se ele, após uma operação de atualização, deixa de estar associado (cf. 3.2) a algum registro de  $A_2$ .

Com esse critério, minimiza-se o número de registros que devem ser alterados pela movimentação de registros na conexão. Existe uma movimentação de um registro do arquivo  $A$  associado a  $b_1$  de  $B$  segundo uma certa conexão quando ele passar a associar-se a  $b_2$  de  $B$  segundo essa conexão, deixando a associação com  $b_1$ ,  $b_1 \neq b_2$ .

4.2.1.2. Se o conjunto de relacionamentos  $L$ , correspondente a  $R$ , é do tipo  $1:n$ , conforme a fig.20,

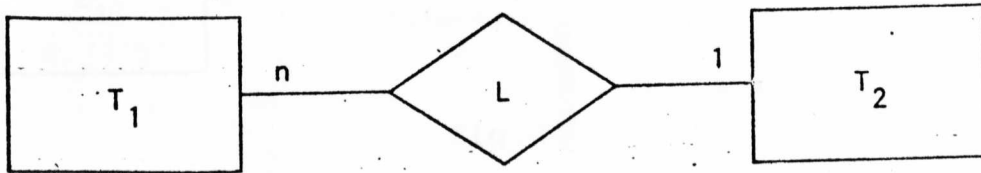


fig.20

- duplicar, em  $E_1$ , a chave de  $E_2$  que havia sido duplicada em  $R$ ; introduzir, em  $E_1$  os atributos de  $R$  correspondentes aos de  $L$ .

- proceder à eliminação (v. 4.2.1.1.) (fig.21).

Observe-se que há uma possibilidade particular de conversão da conexão explícita. Trata-se do caso em que  $L$  é do tipo  $1:x$ , em que  $x \leq 191$  (no sentido do item 2), sem atributos. Nesse caso, poderemos optar por duplicar  $t_1$  em  $E_2$ . Esse novo atributo deve ser declarado com a especificação 'MU'. Existe uma vantagem nessa organização: haverá uma diminuição do tempo de processamento de consultas que solicitam apenas os valores de  $t_1$ , isto é, a chave de  $T_1$ , usando como critério de seleção valores de atributos de  $E_2$ , pois nesse caso o acoplamento não é usado. Por outro lado, existe uma grande desvantagem nessa organização em relação à anterior, sem 'MU': nesta, ocorre um controle automático da in

tegridade, no sentido da preservação do tipo 1:n, enquanto que usando 'MU' essa preservação fica a cargo do usuário.

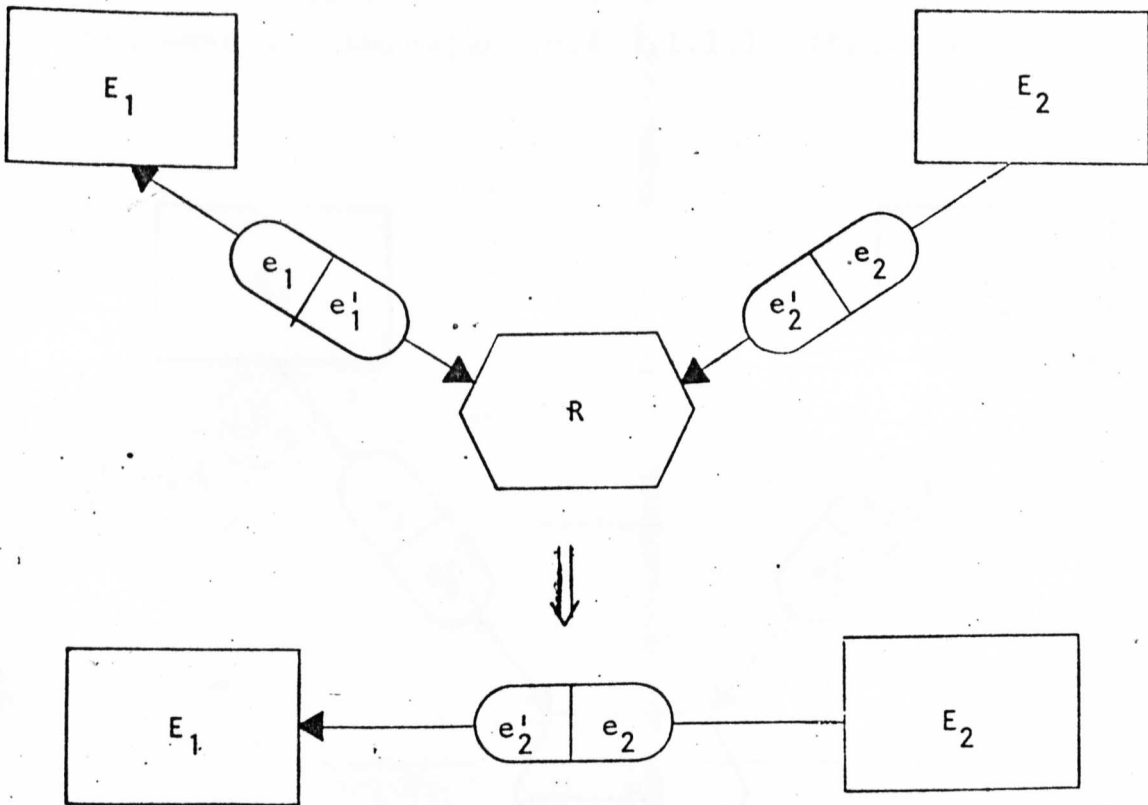


fig.21

4.2.1.3. Se o conjunto de relacionamentos  $L$ , correspondente a  $R$ , é do tipo  $m:y$ , em que  $k < y < 191$ , um atributos conforme a fig.22.

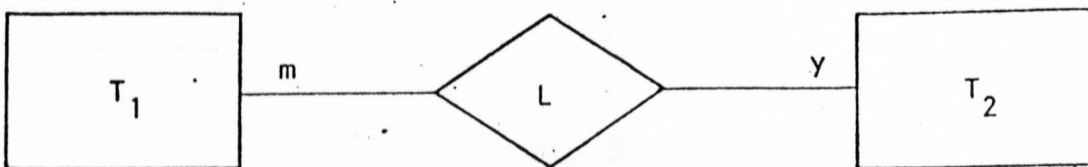


fig.22

- duplicar a chave de  $E_2$  em  $E_1$ , originando um atributo que deve ser especificado como 'MU'.
- proceder à eliminação (v.4.2.1.1.) (fig.23).

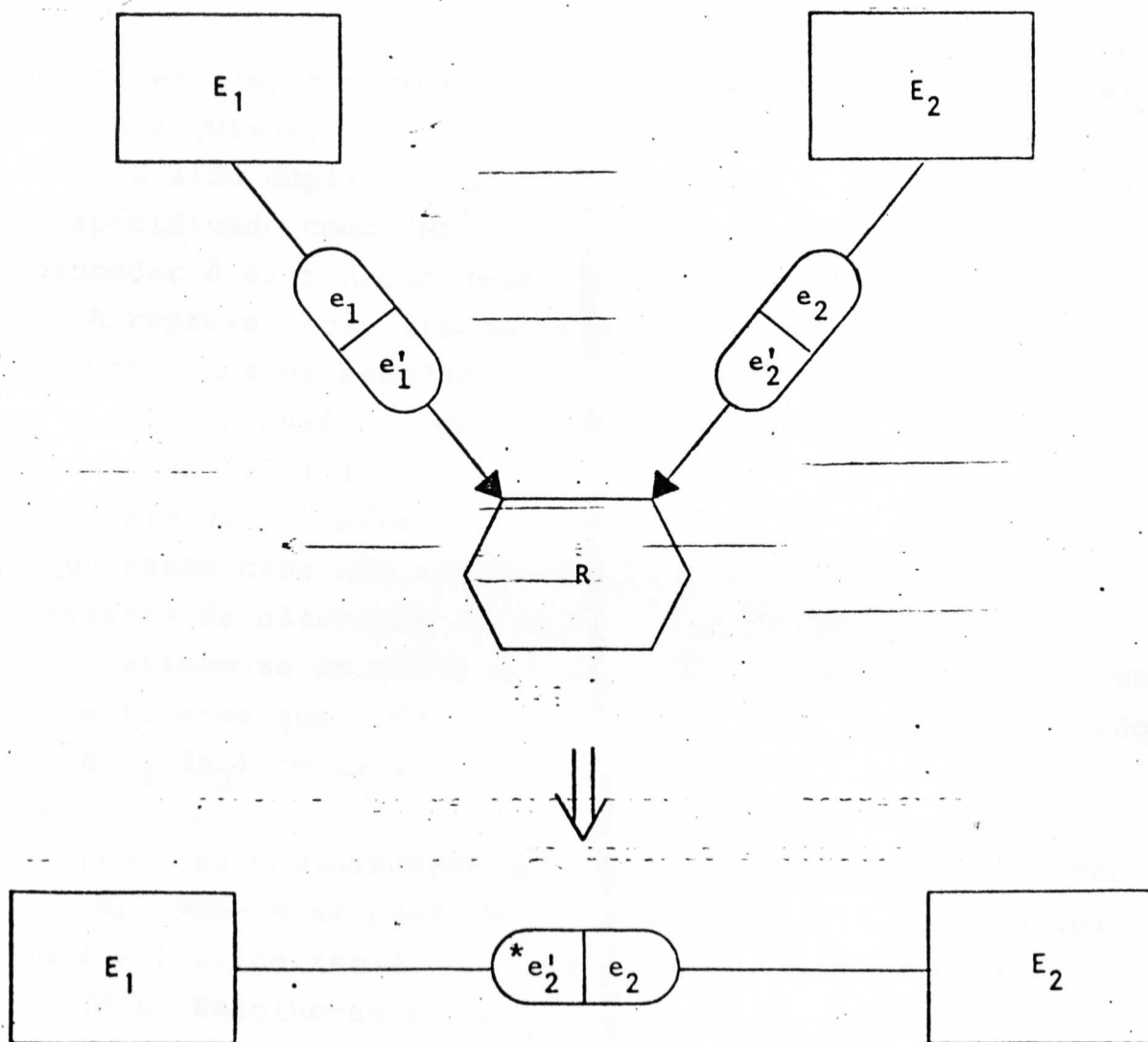


fig.23

4.2.1.4. Se o conjunto de relacionamentos  $L$ , correspondente a  $R$ , é do tipo  $x:y$ , em que  $k \leq x, y \leq l91$ , sem atributos, conforme a fig. 24.

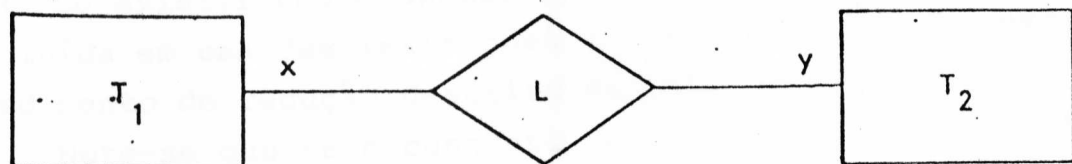


fig.24

- escolher, segundo um ou mais dos critérios descritos abaixo, um dos E-arquivos, suponhamos  $E_1$ , e duplicar nele a chave de  $E_2$  que havia sido duplicada em R, originando um atributo que deve ser especificado como 'MU'.

- proceder à eliminação (v.4.2.1.1.).

A representação diagramática é idêntica à da fig.23.

Critérios de escolha:

- *Critério de ausência no CR*: Idem ao critério do mesmo nome apresentando em 4.2.1.1.

- *Critério da frequência de consultas*: Idem a 4.2.1.1. lembrando que neste caso não existem atributos do CR.

- *Critério de alteração da chave*: Para  $m$  consultas ao banco de dados, estimam-se em média  $m_1$  ( $m_2$ ) atualizações das chaves de  $E_1$  ( $E_2$ ); estima-se que cada registro de  $E_1$  ( $E_2$ ) esteja associado em média a  $n_1$  ( $n_2$ ) registros de  $E_2$  ( $E_1$ ). Escolhe-se  $E_1$  se  $m_1 n_1 > m_2 n_2$ .

- *Critério de movimentação no CR*: Para  $n$  consultas ao banco de dados, estimam-se em média  $n_1$  ( $n_2$ ) movimentações (cf. a definição em 4.2.1.1.) de registros de  $E_1$  ( $E_2$ ) na conexão correspondente ao CR L. Escolhe-se  $E_1$  se  $n_1 > n_2$ .

Esse critério tem efeito análogo ao critério de estabilidade de presença de 4.2.1.1.

- *Critério do comprimento dos registros*: Estima-se que cada registro de  $E_1$  ( $E_2$ ) esteja associado em média a  $n_1$  ( $n_2$ ) registros de  $E_2$  ( $E_1$ ); seja  $s_1$  ( $s_2$ ) o comprimento da chave  $t_1$  ( $t_2$ ). Escolhe-se  $E_1$  se  $n_1 s_2 < n_2 s_1$ .

Com isso, conseguem-se registros com menor tamanho, relativamente à parte variável devido ao grau de multiplicidade do atributo especificado com 'MU'.

4.2.2. Para cada par de E-arquivos,  $E_1$  e  $E_2$ , que estejam relacio

nados entre si por mais de uma conexão explícita binária,

4.2.2.1. Se existir uma e apenas uma dessas conexões que possa ser incluída em uma das restrições 4.2.1.1. a 4.2.1.4., execute o procedimento de redução descrito no item correspondente.

Note-se que se o conjunto de relacionamentos, correspondente a essa conexão, tiver atributos, estes serão implementados em um E-arquivo. Esse fato poderá causar problemas aos outros conjuntos de relacionamentos, que não terão nada a ver com esses atributos.

Por essa razão, poder-se-á eventualmente optar por não executar a redução relativa a essa conexão.

4.2.2.2. Se existir mais de uma dessas conexões que possa ser incluída em uma das restrições 4.2.1.1. a 4.2.1.4., escolha uma delas de acordo com os critérios descritos abaixo e execute o procedimento de redução correspondente.

4.2.3. Para cada par formado um E-arquivo E e um X-arquivo X, relacionados por uma conexão explícita binária estabelecida através de um R-arquivo R e por uma conexão implícita,

4.2.3.1. Se o conjunto de relacionamentos L, correspondente a R, é do tipo 1:1 ou 1:n:

- duplicar em E sua própria chave, duplicando, dessa forma, os relacionamentos que estavam representados em R; introduzir em E os atributos de R correspondentes aos de L.

- os arquivos R e X, bem como os acoplamentos associados, são retirados do modelo.

- *Critério de espaço:* Selecione o CR cujo R-arquivo seja estimado como o que ocupa o maior espaço de arquivamento.

- *Critério de eficiência:* Selecione o CR para o qual se estima que a transformação de sua conexão explícita em implícita produza o maior aumento de eficiência às consultas ao sistema.

Esse critério depende não só dos tipos de consulta, mas também da frequência relativa de cada tipo em relação aos outros tipos.

- *Critério da ausência de atributos:* Escolha um dos CRs que tenha o menor número de atributos.

Com esse critério evitam-se ou diminuem-se os efeitos indesejáveis comentados em 4.2.2.1.

## 5. CONCLUSÕES

Creemos ter mostrado com este trabalho a viabilidade da utilização do MER no projeto de bancos de dados usando pacotes comerciais, tendo para isso empregado o sistema ADABAS com exemplo. É verdade que esse sistema, devido à sua simplicidade, presta-se bem a esquematizações e formalizações. Parece-nos, no entanto, que a mesma metodologia poderia ser desenvolvida para outros pacotes, eventualmente com maior dificuldade.

Por falta de espaço deixamos de apresentar um exemplo de aplicação, que pode ser obtido com os autores.

Um tópico a ser desenvolvido em detalhe seria o de mostrar que o procedimento descrito produz arquivos em ADABAS que, se vistos como relações do Modelo Relacional /COD 70/, estão em geral em 3ª Forma Normal /COD 72/. Este não é o caso das conexões implícitas que empregam a especificação 'MU', pois aí temos relações que não estão nem em 1ª Forma Normal. Com isso, a metodologia aqui descrita corresponde, praticamente, a um projeto "Top-Down" de relações normalizadas implementadas em ADABAS. A correspondência do ADABAS com o Modelo Relacional foi abordada por /REN 76/ sem que, no entanto, fosse dada uma metodologia de estruturação do modelo.

## REFERÊNCIAS

- /BAC 69/ Bachman, C.W. - "Data Structure Diagrams", *Data base* (journal of the ACM SIGEDP) 1,2, 1969, pp.4-10.
- /CHE 76/ Chen, P.P.-S. "The Entity-Relationship Model - Toward a Unified View of Data", *ACM Transactions on Data Base Systems* 1,1, Mar. 1976, pp.9-36.
- /CHE 77/ Chen, P.P.-S. "The Entity-Relationship Model - a basis for the enterprise view of data", *Proc. AFIPS 1977 NCC*, Vol.46, AFIPS Press, Montvale, pp.77-84.
- /COD 70/ Codd, E.F. "A Relational Model of Data for Large Shared Data Banks", *Comm. ACM* 13, June 1970, pp.377-387.

- /COD 72/ Codd, E.F. "Further Normalization of the Data Base Rela  
tional Model", in Rustin, R. (ed.) *Data Base Systems*,  
Courant Computer Science Symposium 6, Prentice-Hall,  
Englewood Cliffs, 1972.
- /DAT 77/ Date, C.J. *An Introduction to Database Systems* (2<sup>nd</sup>  
ed.), Addison Wesley, Reading, 1977.
- /MRI 72/ SYS 2000 Publications: General Information Manual,  
Basic Reference Manual, MRI Systems Corp., Austin, 1972.
- /REN 76/ Renaud, D. "ADABAS and the Relational Model" 3<sup>rd</sup> Inter-  
national ADABAS User's Conference, San Francisco, May  
76, pp.177-185.
- /S-N 79/ Santos, C.S., Neuhold, E.J., Furtado, A.L. "A data type  
approach to the Entity-Relationship Model", PUC, Rio de  
Janeiro, 1979.
- /SOF 76/ ADABAS Publications: ADABAS Introduction, Mar. 1979;  
ADASCRIP+ Reference Manual, Mar. 1979, ADABAS Ref.  
Manual, Sept. 1976, NATURAL Ref. Manual, Jan. 1980,  
Software AG, Darmstadt.
- /S-S 79/ Schiffner, G., Scheuermann, P. "Multiple views and  
abstractions with an Extended-Entity-Relationship Model"  
Computer Languages 4, 1979, pp.139-154.

"RELATÓRIO TÉCNICO"  
DEPARTAMENTO DE MATEMÁTICA APLICADA  
TÍTULOS PUBLICADOS

- RI-MAP-7701 - Ivan de Queiroz Barros  
On equivalence and reducibility of Generating Matrices of RK-Procedures - Agosto 1977
- RI-MAP-7702 - V.W. Setzer  
A Note on a Recursive Top-Down Analyzer of N.Wirth - Dezembro 1977
- RI-MAP-7703 - Ivan de Queiroz Barros  
Introdução a Aproximação Ótima - Dezembro 1977
- RI-MAP-7704 - V.W. Setzer, M.M. Sanches  
A linguagem "LEAL" para Ensino básico de Computação - Dezembro 1977
- RI-MAP-7801 - Ivan de Queiroz Barros  
Proof of two Lemmas of interest in connection with discretization of Ordinary Differential Equations - Janeiro 1978
- RI-MAP-7802 - Silvio Ursic, Cyro Patarra  
Exact solution of Systems of Linear Equations with Iterative Methods  
Fevereiro 1978
- RI-MAP-7803 - Martin Grötschel, Yoshiko Wakabayashi  
Hypohamiltonian Digraphs - Março 1978
- RI-MAP-7804 - Martin Grötschel, Yoshiko Wakabayashi  
Hypotractable Digraphs - Maio 1978
- RI-MAP-7805 - W. Hesse, V.W. Setzer  
The Line-Justifier: an example of program development by transformations  
Junho 1978
- RI-MAP-7806 - Ivan de Queiroz Barros  
Discretização  
Capítulo I - Tópicos Introdutórios  
Capítulo II - Discretização  
Julho 1978
- RI-MAP-7807 - Ivan de Queiroz Barros  
(1', 1') - Estabilidade e Métodos Preditores-Corretores - Setembro 1978
- RI-MAP-7808 - Ivan de Queiroz Barros  
Discretização  
Capítulo III - Métodos de passo progressivo para Eq. Dif. Ord. com condições iniciais - Setembro 1978
- RI-MAP-7809 - V.W. Setzer  
Program development by transformations applied to relational Data-Base Queries - Novembro 1978
- RI-MAP-7810 - Nguiffo S. Boyom, Paulo Boulos  
Homogeneity of Cartan-Killing spheres and singularities of vector fields - Novembro 1978

TÍTULOS PUBLICADOS

- RT-MAP-7811 - D.T. Fernandes e C. Patarra  
Sistemas Lineares Esparsos, um Método Exato de Solução - Novembro 1977
- RT-MAP-7812 - V.W. Setzer e C. Bressan  
Desenvolvimento de Programas por Transformações: uma Comparação entre dois Métodos - Novembro 1978
- RT-MAP-7813 - Ivan de Queiroz Barros  
Varição do Passo na Discretização de Eq. Dif. Ord. com Condições Iniciais - Novembro 1978
- RT-MAP-7814 - Martin Grötschel e Yoshiko Wakabayashi  
On the Complexity of the Monotone Asymmetric Travelling Salesman Polytope I: HYPHAMILTONIAN FACETS - Dezembro 1978
- RT-MAP-7815 - Ana F. Humes e E.I. Jury  
Stability of Multidimensional Discrete Systems: State-Space Representation Approach - Dezembro 1978
- RT-MAP-7901 - Martin Grötschel, Yoshiko Wakabayashi  
On the complexity of the Monotone Asymmetric Travelling Salesman Polytope II: HYPOTRACEABLE FACETS - Fevereiro 1979
- RT-MAP-7902 - M.M. Sanches e V.W. Setzer  
A portabilidade do Compilador para a Linguagem LEAL - Junho 1979
- RT-MAP-7903 - Martin Grötschel, Carsten Thomassen, Yoshiko Wakabayashi  
Hypotractable Digraphs - Julho 1979
- RT-MAP-7904 - N'Guiffo B. Boyom  
Translations non triviales dans les groupes (transitifs) des transformations affines - Novembro 1979
- RT-MAP-8001 - Ângelo Barone Netto  
Extremos detectáveis por jatos - Junho 1980
- RT-MAP-8002 - Ivan de Queiroz Barros  
Medida e Integração  
Cap. I - Medida e Integração Abstrata - Julho 1980
- RT-MAP-8003 - Routo Terada  
Fast Algorithms for NP-Hard Problems which are Optimal or Near-Optimal with Probability one - Setembro 1980
- RT-MAP-8004 - V.W. Setzer e R. Lapyda  
Uma Metodologia de Projeto de Bancos de Dados para o Sistema ADARAS  
Setembro 1980
- RT-MAP-8005 - Imre Simon  
On Brzozowski's Problem: (LUA)<sup>m</sup> = A\* - Outubro 1980
- RT-MAP-8006 - Ivan de Queiroz Barros  
Medida e Integração  
Cap. II - Espaços Lp - Outubro 1980

TÍTULOS PUBLICADOS

- RT-MAP-8101 - Lúcia Kazuko Yoshida e Gabriel Richard Bitran  
Um algoritmo para Problemas de Programação Vetorial com Variáveis  
Zero-um - Fevereiro 1981
- RT-MAP-8102 - Ivan de Queiroz Barros  
Medida e Integração  
Cap. III - Medidas em Espaços Topológicos - Março 1981
- RT-MAP-8103 - V.N. Setzer, R. Lepyda  
Design of Data Models for the ADAMS System using the Entity-Relationship  
Approach - Abril 1981
- RT-MAP-8104 - Ivan de Queiroz Barros  
Medida e Integração  
Cap. IV - Medida e Integração Vetoriais - Abril 1981
- RT-MAP-8105 - U.S.R. Murty  
Projective Geometries and Their Truncations - Maio 1981
- RT-MAP-8106 - V.N. Setzer, R. Lepyda  
Projeto de Bancos de Dados Usando Modelos Conceituais  
Este relatório técnico complementa o RT-MAP-8103. Ambos substituem o  
RT-MAP-8064 ampliando os conceitos ali expostos. - Junho 1981
- RT-MAP-8107 - Maria Angela Gursel, Yoshiko Wakabayashi  
Embedding of Trees - August 1981
- RT-MAP-8108 - Ivan de Queiroz Barros  
Mecânica Analítica Clássica - Outubro 1981