

# 5

## Bases de dados

### OBJETIVOS

Em sentido lato, bases de dados são o modo como os dados são armazenados em computadores. Elas podem ser:

- uma coleção de dados sobre as atividades de uma organização, que assim permite o controle dessas atividades
- uma coleção de dados disponíveis publicamente, mantidos num computador hospedeiro ou servidor acessível por meio de rede de telecomunicações ou em cederrom.

Ao término deste capítulo você:

- estará conhecendo os conceitos de registros e campos
- poderá reconhecer os diferentes tipos de bases de dados que podem ser importantes para os gerentes da informação
- terá compreendido a natureza do formato de registro MARC para bases de dados bibliográficos
- terá compreendido a diferença entre estruturas físicas e lógicas de bases de dados
- terá conhecido a necessidade das estruturas de bases de dados.

### 5.1 Introdução

Os gerentes de bibliotecas e serviços de informação sempre compilaram arquivos de informações, fossem catálogos ou listas de leitores registrados. Os antigos sistemas informatizados de muitas empresas conservavam arquivos-mestre, que em geral continham dados relativos a folha de pagamento, vendas, compras e inventário. Essas aplicações compreendem uma série de registros relacionados entre si e formatados de modo semelhante. Bases de dados externas podem ser acessadas por intermédio dos serviços de buscas em linha ou adquiridas em cederrom. O gerente de informação poderá importar partes dessas bases de dados, desde que disponha das licenças necessárias, a fim de integrá-las em bases de dados locais. Em virtude de as bases de dados serem essenciais para a forma como os dados são armazenados e recuperados, é

importante que o gerente de informação conheça os tipos de bases de dados que existem, quaisquer formatos-padrão de registros que sejam encontrados, bem como os métodos de estruturação de bases de dados.

Este capítulo, portanto, apresenta algumas definições básicas relativas a bases de dados e depois examina os tipos de bases de dados. Em seguida, trata dos formatos-padrão de registro, e a parte final explica algumas idéias fundamentais sobre estrutura de bases de dados.

### 5.2 Definições básicas

**Arquivo** é uma coleção de registros similares, com relações definidas entre si. Um sistema de base de dados pode compreender vários arquivos interligados. Examinamos isso mais detidamente na seção 5.5 que trata das estruturas de bases de dados. Para dar início à nossa análise focalizaremos a base de dados ou arquivo tomado isoladamente.

**Registro** é a informação contida na base de dados e que diz respeito a um documento ou item. Por exemplo, numa base de dados catalográficos um registro conterá todas as informações acerca de determinado livro. Numa base de dados de fontes, um registro mostrará o conteúdo do verbete de um cadastro, ou artigo de periódico. Numa base de dados de inventário, o registro conterá todas as informações pertinentes quanto ao estoque de determinado item. Os registros compõem-se de vários campos. Os tipos de campos utilizados, sua extensão e a quantidade de campos de um registro são escolhidos de acordo com a aplicação específica. A figura 5.1 relaciona os tipos de campos que se encontram no arquivo de um sistema de controle de inventário, e a figura 5.2 mostra os campos que se encontram num registro bibliográfico.

N.º do item
Saldo
Nível para fazer nova encomenda
Nível de perigo
Preço de venda
Descrição do item
Localização
Quantidade disponível
Código do fornecedor do item
Fornecedor do item
Unidade de encomenda
Número/unidade
Último custo
Preço unitário de venda
Valor do estoque atual pelo último custo

Figura 5.1

Alguns campos comuns de um arquivo-mestre de inventário

```

USMARC record
001 $a94-790547
005 $a19950106130304.3
008 950106s1994###dun#####eng#
040 $aDLC:$dLC$aDLC
050 00$aZ695.615
082 105$a025.3:$212
111 25$aSeminar on Cataloging Digital Documents $d(1994) : $cUniversity of Virginia
    Library and Library of Congress)
245 105$aProceedings of the Seminar on Cataloging Digital Documents, October 12-
    14, 1994 $h[computer file]
    /$cUniversity of Virginia Library, Charlottesville, and the Library of
    Congress.
256 $aComputer data and program.
260 $a(Washington, D.C. : $bLibrary of Congress, $c1994).
538 $aAccess: Internet. Address:
    http://lcwebloc.gov/catdir/senddigdocs/seminar.html
500 $aTitle from title screen.
500 $a Sponsor: Sarah Thomas, director for cataloging, Library of Congress" -- Home
    page.
520 $aText, graphics, and audio files, including a summary of the seminar by Sarah
    Thomas, color photographs of the presenters and various events, texts of the
    presentations, notes taken by Library of Congress staff, records of the panel
    discussion, an action plan, and a list of participants.
650 #0$aCataloging of computer files $xCongresses $xDatabases.
700 1 # $aThomas, Sarah.
710 2 # $aUniversity of Virginia. $bLibrary.
710 2 # $aLibrary of Congress.
856 7 $uhttp://lcwebloc.gov/catdir/senddigdocs/seminar.html#2http

```

**Figura 5.2** Campos de um registro bibliográfico (USMARC). *ver também seção 5.4)*

Existem dois tipos de campos: campos de tamanho fixo e campos de tamanho variável. Um **campo de tamanho fixo** contém o mesmo número de caracteres em cada registro. Como os tamanhos fixos são previsíveis, não é preciso sinalizar para o computador onde começa e termina cada campo. Os campos de tamanho fixo permitem um armazenamento econômico e os registros que os adotam podem ser codificados rápida e facilmente. No entanto, os campos de tamanho fixo talvez não comportem de forma adequada dados de tamanho variável. Os campos de tamanho fixo são excelentes para serem usados com códigos, como ISBNs, códigos de leitores, códigos de produtos, números de contas bancárias, datas e códigos de idiomas, cuja extensão da informação será igual em cada registro. Quando se trata de dados de tamanho variável tornam-se necessários campos de tamanho variável. Um **campo de tamanho variável** terá diferentes tamanhos em diferentes registros. Neste caso, o computador não pode reconhecer quando um campo termina e outro começa, tornando-se necessário, portanto, sinalizar o início e o fim dos campos. A análise do formato de registro MARC trata de uma das formas de se realizar isso. As bases de dados

estudadas neste capítulo empregam uma mistura de campos fixos e variáveis, a fim de poderem comportar os diferentes tipos de dados. Além disso, objetos como fotografias e vídeos podem ser armazenados como arquivos separados ligados a registros que contenham campos de tamanho essencialmente fixo ou variável.

### Reflexão

**Examine os campos do registro bibliográfico da figura 5.2. Por que cada um deles foi incluído? Quais desses itens de dados poderiam ser acomodados satisfatoriamente em campos de tamanho fixo?**

Dentro de cada campo, os elementos específicos de dados ou unidades de informação podem ser denominados **subcampos**. Estes subcampos precisam ser sinalizados, a fim de que possam ser identificados pelo computador.

Os registros de uma base de dados podem ser recuperados por meio da inserção de séries de caracteres e solicitando-se ao computador que as localize. Isso permitirá que sejam selecionados subconjuntos da base de dados conforme um critério de consulta. Evidentemente, o que se pode recuperar depende do que existe na base de dados e da forma como as informações tiverem sido estruturadas. Comumente, são feitas buscas em certas partes dos registros e seu conteúdo é usado como **chaves de busca**. É comum haver uma gama razoável de tipos de chaves de busca, bem como diferentes maneiras pelas quais se pode fazer uma busca numa base de dados. Também é possível, usualmente, selecionar os campos de um registro que são apresentados numa determinada aplicação e formatar essa apresentação de várias maneiras diferentes.

### 5.3 Tipos de bases de dados

As bases de dados podem ser armazenadas em meios magnéticos ou ópticos, como discos, e acessadas local ou remotamente. Isso inclui o acesso à base de dados da instituição, que abranja transações e registros financeiros, ou a outras bases de dados que possam ser acessadas à distância. Algumas dessas bases de dados conterão informações de acesso público, como resumos e índices, textos completos de relatórios, enciclopédias e cadastros, enquanto outras serão compartilhadas no âmbito de uma instituição ou grupo de instituições.

### Bases de dados de referências e de fontes

Classificam-se como de referências e de fontes as bases de dados disponíveis para os usuários da informação em âmbito público e que sejam acessadas à distância, por meio de um serviço de buscas em linha, ou localmente em cederrrom.

As **bases de dados de referências** remetem ou encaminham o usuário a outra fonte, como um documento, uma pessoa jurídica ou pessoa física, para que

obtenha informações adicionais, ou o texto completo de um documento. Como exemplos temos:

- 1 **Bases de dados bibliográficos**, que incluem citações ou referências bibliográficas e, às vezes, resumos de trabalhos publicados. Informam ao usuário sobre o que foi publicado e onde se publicou (por exemplo, se num periódico, nos anais de um congresso) e, na hipótese de a base conter resumos, apresentarão uma síntese do conteúdo do documento original.
- 2 **Bases de dados catalográficos**, que mostram o acervo de determinada biblioteca ou rede de bibliotecas. Comumente, essas bases de dados relacionam quais os livros, títulos de periódicos e outros itens que a biblioteca possui em seu acervo, porém não proporcionam informações adicionais sobre o conteúdo desses documentos. As bases de dados catalográficos são um tipo especial de base de dados bibliográficos, porém, em virtude de sua orientação ser bastante diferente da de outras bases de dados bibliográficos, merecem ser identificadas como uma categoria à parte.
- 3 **Bases de dados referenciais**, que referenciam informações ou dados, como nomes e endereços de instituições, e outros dados típicos de cadastros.

**Reflexão** Examine um dos registros encontrados no catálogo da biblioteca que lhe seja mais acessível. Quais os campos que ele apresenta e que podem ser visualizados por você como usuário?

As bases de dados de fontes contêm os dados originais e constituem um tipo de documento eletrônico. Após ter feito uma consulta bem-sucedida numa base de dados de fontes, o usuário terá em mãos as informações de que precisa, sem ter de ir buscá-las numa fonte original (como seria obrigado a fazer no caso das bases de dados de referências). Os dados encontram-se disponíveis tanto em formato legível por computador, quanto em formato impresso. As bases de dados de fontes podem ser agrupadas, segundo seu conteúdo, em:

- 1 Bases de dados numéricos, que contêm dados numéricos de vários tipos, inclusive dados estatísticos e de levantamentos.
- 2 Bases de dados de texto integral, que contêm notícias de jornal, especificações técnicas e programas de computador.
- 3 Bases de dados textuais e numéricos, que contêm uma mistura de dados textuais e numéricos (como, por exemplo, relatórios anuais de empresas) e dados de manuais.
- 4 Bases de dados multimídia, que incluem informações armazenadas numa mescla de diferentes tipos de meios, inclusive, por exemplo, som, vídeo, fotografias, textos e animação.

As bases de dados bibliográficos contêm uma série de registros bibliográfico, ligados entre si, onde cada um em geral apresenta uma combinação dos seguintes componentes:

- número do documento
- título
- autor
- referência da fonte
- resumos
- texto integral
- termos ou expressões de indexação
- citações ou quantidade de referências
- instituição de origem do documento, ou endereço do autor, ou ambos
- língua do documento-fonte
- informação de uso interno, como números de classificação ou localização.

Esses componentes podem ser denominados como uma referência ou citação documental, e cada um deles geralmente será representado por um campo diferente. Há notável variabilidade entre os formatos de registro de bases de dados bibliográficos, de modo que a lista acima menciona apenas os componentes mais comuns, que não fornecem as informações ou o texto do documento-fonte, mas somente indicam onde as informações podem ser encontradas. É possível também encontrar em muitas bases de dados um resumo junto a cada referência, que, se for um bom resumo informativo, proporcionará, imediatamente, informações úteis, mesmo se tratando de uma base de dados bibliográficos. Empregam-se alguns desses componentes, mais comumente, como **chaves de recuperação primárias** (por exemplo, nome do autor, palavras do título, termos de indexação), enquanto outros serão mais úteis como **chaves de recuperação secundárias** (por exemplo, para limitar o conjunto de registros recuperados numa busca sobre um termo de indexação temática). As chaves de recuperação secundárias usuais são: língua, informações de uso interno, título de periódico. Os elementos restantes do registro só são exibidos no vídeo ou impressos para oferecer ao usuário informações adicionais sobre o documento, sendo úteis na avaliação da relevância ou na localização do documento-fonte.

Embora, há mais de vinte anos, haja inúmeras bases de dados bibliográficos informatizadas, seus elementos básicos ainda retêm a marca de sua origem no produto impresso, o qual, geralmente, era uma publicação de resumos ou um índice aos quais essas bases estavam vinculadas. Muitas vezes foram originalmente construídas para servir de apoio à produção mais eficiente de um serviço de resumos ou índice impresso, e, com frequência, é ainda o produto impresso que responde por uma parcela significativa da receita do produtor da base de dados. Uma das vantagens da manutenção de registros em formato legível por computador era a possibilidade assim criada de se gerar uma série de produtos

**Índices multidisciplinares**  
*Science Citation Index*  
*Social Sciences Citation Index*  
*Arts and Humanities Citation Index*  
 Disponíveis nos formatos: impressos, acumulações impressas, discos compactos, discos compactos com resumos, fitas magnéticas, em linha, Intranet

**CompuMath Citation Index**  
 Disponível nos formatos: impresso, acumulação impressa, em linha, cederrrom

**Índices específicos de uma disciplina em disco compacto**  
*Biochemistry and Biophysics Citation Index*  
*Biotechnology Citation Index*  
*Chemistry Citation Index*  
*Neuroscience Citation Index*  
*Materials Science Citation Index*

**Índices do conteúdo de livros, anais e resenhas**  
*Index to Scientific Book Contents*  
*Index to Scientific and Technical Proceedings*  
*Index to Social Sciences and Humanities Proceedings*  
*Index to Scientific Reviews*

**Serviços por encomenda que incluem:**  
*Research Alert*  
*Medical Documentation Service*  
*Contract Research*  
*Science Watch*

**Produtos na área da química que incluem:**  
*Current Chemical Reactions* (impresso)  
*Current Chemical Reactions In-House Database* (fitas magnéticas, acumulação em fitas magnéticas)  
*Index Chemicus* (impresso e em cederrrom)  
*ChemPrep on CD-ROM*

**Fitas magnéticas que incluem:**  
*Current Contents Search*  
*SciSearch*  
*Social SciSearch*  
*Arts and Humanities Search*  
*Research Alert*

**Bases de dados em linha que incluem:**  
*Current Contents Search*

*Continua ao lado*

**Figura 5.3** Produtos de um grupo de bases de dados afins de um produtor: o Institute for Scientific Information (ISI)

*SciSearch*  
*Social SciSearch*  
*Arts and Humanities Search*  
*Computer and Mathematics Search*  
*ISTP Search*  
*ISTP&B Search*

**Sumários correntes (*Current contents*) em diversos campos, por exemplo, ciências da vida; agricultura; biologia e ciências ambientais**  
 Impressos, disquetes, cederrroms, em linha, fitas magnéticas, Internet

**Outros produtos que incluem:**  
*The Genuine Article* (fornecimento de documentos)  
*ISI Document Solution*  
*Request-a-Print* (cartões impressos de pedidos de separatas)  
*Pro-Cite e Reference Manager* (pacotes de programas bibliográficos)  
*Journal Citation Reports*  
*Social SciSearch*  
*Arts and Humanities Search*  
*Research Alert*

**Figura 5.3** Continuação

**Tabela 5.1** Algumas bases de dados bibliográficos do SIN International

Base de dados	Produtor	Conteúdo
ABI-INFORM	University Microfilms Inc.	Negócios e administração
BIOSIS Previews	BIOSIS	Bio-ciências
CA	CAS (Chemical Abstracts)	Química
COMPENDEX	Engineering Information Inc.	Engenharia
DISSABS	University Microfilms Inc.	Bibliográfica especializada em teses da América do Norte
EMBASE	Elsevier Science Publishers	Biomedicina e fármacos
EVENTINE	Elsevier Science Publishers	Congressos e eventos multidisciplinares
INPADOC	European Patent Office	Patentes internacionais
INSPEC	Institution of Electrical Engineers	Engenharia, física e eletrônica
JGRIP	The Japan Science and Technology Corporation	Pesquisas multidisciplinares de instituições públicas do Japão
MEDLINE	us National Library of Medicine	Medicina
SCISEARCH	Institute for Scientific Information	Ciência e tecnologia, bibliográfica de citações
WORDCAT	OCLC Online Computer Library Center	Multidisciplinar, com os catálogos integrados de bibliotecas de vários países

Tabela 5.2 Algumas bases de dados de fontes do Oeste/Orbis

Base de dados	Produtor	Conteúdo
EDOC	European Patent Office	Aplicações de patentes
FMARK	Institut National de la Propriété Industrielle	Marcas francesas
DUNS United Kingdom	Dun & Bradstreet	Informações sobre empresas do Reino Unido
ASPO	Agence France-Presse	Notícias esportivas francesas e internacionais, reportagens, resultados, resumos e biografias na área esportiva
LOGOS	Documentation Française	Informação com texto integral sobre política, sociedade e economia da França
PROMT	Information Access Company	Artigos sobre produtos, mercados e tecnologias
Who's Who in Technology	Gale Research Inc.	Informações sobre figuras de destaque da tecnologia norte-americana
Gale Directory of Databases	Gale Research Inc.	Descrições de bases de dados disponíveis em cederrom, em linha e outros formatos
Chemical Economics Handbook	SRI International	Dados de pesquisa de mercado para a indústria química

diferentes para diferentes mercados a partir de um único conjunto de registros (ver figura 5.3). Tais produtos incluem amplitude serviços de notificação corrente e de buscas em linha, concessão de licenças para importação de partes de bases de dados, e serviços impressos de resumos e índices.

As bases de dados de fontes são documentos eletrônicos. A tabela 5.2 mostra alguns exemplos. O conteúdo dessas bases de dados pode ser tão variado quanto o conteúdo de um livro impresso, e pode ainda incluir textos, números, tabelas, figuras e gráficos. Na realidade, muitas dessas bases de dados aproveitaram-se do fato de não se acharem restringidas pelas mesmas limitações físicas dos materiais impressos, sendo multimiídia, pois incluem, além de textos e dados numéricos, programas de computador, imagens, som, mapas e gráficos. Elas podem ser acessadas em linha por meio de serviços de buscas em linha, em cederroms ou por meio de videotexto e teletexto. O cederrom é o meio que suportou as realizações iniciais das mais ambiciosas bases de dados multimiídia, pois não é preciso transmitir imagens por meio de redes externas de comunicação.

As bases de dados de fontes são tão variadas quanto à sua natureza e origens que se torna difícil fazer generalizações. Anteriormente, neste capítulo, dividimos as bases de dados de fontes em numéricas, de texto integral, e textuais e

Título: Primary ovarian non-Hodgkin's lymphoma: Outcome after treatment with combination chemotherapy.

Author(s): Dimopoulos, M.A. {a,b,c}; Daliani, D. {d,e,f}; Pugh, W. {g,h}; Gershenson, D. {i,k,l}; Cabanillas, F. {m,n,o}; Sarris, A.H.

Corporate Source: {a} Section Lymphoma, Box 68, Univ. Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA; {d} Dep. Exp. Clin. Oncol., Aarhus Univ. Hosp., Norrebrogade 44, DK-8000 Aarhus, Denmark; {g} Div. Neurosurgery, UCLA Med. Cent., Los Angeles, Ca 90024, USA; {j} H.H. McGuire VAMC, Dep. Radiol. (114), 1201 Broad Rock Blvd., Richmond, Va 23249, USA; {m} Dep. Otorhinolaryngology, Univ. Tampere Med. Sch., PO Box 607, FIN-33101 Tampere, Finland {b} http://www.utah.edu/lymphoma/~mmdimopoulos; {e} http://www.aarhushosp.ed/eco/~ddaliani; {h} http://www.uclamed.edu/neurosurg/~wpugh; {k} http://www.hnmvami.edu/radiology/~dgershenson; {n} http://www.utampmededu/otorhm/~fcabanillas {c} mmdimopoulos@utmed.edu; {f} ddaliani@aarhushosp.edu; {i} wpugh@uclamed.edu; {l} dgershenson@hnmvamc.edu; {o} fcabanillas@utampmed.edu

Source: Gynecologic Oncology 64(3) [cited August 21, 1997], 1997, 446-450.

http://www.gynecology.com/

Document Type: Article: Research Article

ISSN: 0090-8258

Language: English (EN)

Abstract: Because the outcome of patients with primary ovarian non-Hodgkin's lymphoma (NHL) is controversial, we retrospectively analyzed experience with adults seen at the University of Texas M. D. Anderson Cancer Center from 1974 to 1993. Patients were included if at least one ovary was pathologically involved, and if combination chemotherapy was used that must have included doxorubicin for intermediate grade histologies. We identified 15 patients who constituted 0.5% of all untreated NHL and 1.5% of untreated ovarian neoplasms that presented to our institution during this time. One patient refused therapy, leaving 14 assessable for response. Nine patients had intermediate-grade, 5 had high grade, and none had low-grade NHL. One ovary was involved in 4 patients, and both in 10, in 7 of whom additional sites were involved, including supradiaphragmatic nodes in 2. Four patients had AAS1 and 10 had AAS IV. Favorable (0 or 1) and unfavorable (gt 1) IPI scores were seen in 5 and 9 patients, respectively. The complete remission rate for all patients was 64%, and 5-year survival and FFS for all assessable patients were 57 and 46%, respectively. We conclude that the complete remission rate and FFS of patients with ovarian NHL treated with appropriate chemotherapy appear to be similar to that of patients with other nodal NHLs. Further work is required to determine prognostic factors in ovarian NHL.

Major Concepts: Oncology; Human Medicine, Medical Sciences; Pharmacology

Super Taxa: Hominiidae; Primates, Mammalia, Vertebrata, Chordata, Animalia

Organisms: human (Hominiidae); adult, female, patient

Taxa Notes: animals, chordates; humans; mammals; primates; vertebrates

Diseases: primary ovarian non-Hodgkin's lymphoma; blood and lymphatic disease, combination chemotherapy, endocrine disease/gonads, neoplastic diseases outcome; reproductive system disease/female; malarial infection

Chemicals & Biochemicals: bleomycin; antineoplastic-drug; combination therapy; cyclophosphamide; antineoplastic-drug; combination therapy; dexamethasone;

Figura 5.4 Registros de uma base de dados

antineoplastic-drug; combination therapy; doxorubicin; antineoplastic-drug; combination therapy; etoposide; antineoplastic-drug; combination therapy; methotrexate; antineoplastic-drug; combination therapy; prednisone; antineoplastic-drug; combination therapy; vincristine; antineoplastic-drug; combination therapy; Dithranol; Iron(II)

Registry Numbers: 9041-93-4; PUBMEDCN: 1143-38-0Q; PUBRANOL; 15438-31-0

Methods & Equipments: radiotherapy; therapeutic method

Alternate Indexing: MeSH: Lymphoma, Non-Hodgkin's Ovarian Neoplasms

Notes: This Journal has 16 issues per volume.

Record Type: Abstract

Title: Survival rates and their relationship to life-history traits in some common British birds.

Author(s): Dobson, A. {a}

Corporate Source: {a} Biology Dep., Univ. Rochester, Rochester, NY 14627, USA

Source: Current Ornithology, E. Power, D.M.: Ed.7 1990, 115-146

Document Type: Book Chapter; Review

ISSN: 0742-390X

ISBN: 0-306-43307-9

Book Publisher: Plenum Press, New York, New York, USA; London, England, UK

Language: English (EN)

Major Concepts: Ecology: Environmental Science; Mathematical Biology; Models and Simulations; Computational Biology; Morphology; Physiology; Reproductive System; Reproduction; Systematics and Taxonomy

Super Taxa: Aves - Unspecified; Aves; Vertebrata, Chordata, Animalia

Organisms: bird (Aves); Aves (Vertebrata)

Taxa Notes: animals, birds; chordates; nonhuman vertebrates; vertebrates

Geopolitical Locations: UK; Europe, Palearctic region

Miscellaneous Descriptors: adult; banding data; body weight; Britain; fecundity; haldane's mathematical model; lack's mathematical model; life-history trait; migratory strategy; population studies; survival rate

Figura 5.4 Continuação

numéricas. Nesse contexto, também temos de contemplar as bases de dados referenciais. Estas, embora possam ser classificadas como bases de dados de referências, no sentido de que oferecem uma indicação para se chegar a informações adicionais, freqüentemente são também bases de dados de fontes, pois podem conter o texto integral de um cadastro que fosse considerado como um documento-fonte. As bases de dados de fontes, portanto, podem incluir o texto integral de artigos de periódicos, boletins, noticiários, dicionários, cadastros e outros materiais-fonte. Muitas, mas não todas, possuem equivalente impresso. Algumas não abrangem o conteúdo completo de seu equivalente impresso, pois oferecem somente uma cobertura seletiva.

#### 5.4 Formatos de registro bibliográfico

Todos os registros num arquivo possuem um formato-padrão. A fim de facilitar o intercâmbio de registros entre diferentes sistemas informatizados, foram

encetados esforços visando a desenvolver alguns formatos padronizados de registro. Esses formatos seriam especialmente úteis em aplicações catalográficas, onde um formato-padrão, que também incorporasse um acordo quanto aos elementos do registro bibliográfico, seria particularmente atraente por permitir o intercâmbio de registros catalográficos. Esse intercâmbio diminuiria a necessidade de catalogação nas próprias bibliotecas. Por conseguinte, uma das áreas em que o formato-padrão de registro mais se consolidou foi a da criação de registros catalográficos. O trabalho realizado nesse campo levou ao surgimento do formato de registro MARC.

#### O formato de registro MARC

O formato de registro MARC emprega um formato de diretório para lidar com campos de tamanho variável e de tamanho fixo. A posição do caractere no início de cada campo é especificada por um diretório no começo do registro.

O formato de registro MARC foi desenvolvido pela Library of Congress e a British Library para a comunicação de descrições bibliográficas em formato legível por computador, de modo que os registros pudessem ser reformataados para atender a qualquer objetivo imaginável. Os primeiros ensaios realizados, por volta de 1966, pela Library of Congress, utilizaram o formato MARC I; este formato, porém, foi posto de lado em 1967, sendo substituído pelo MARC II ou MARC, como é denominado comumente.

À medida que um maior número de países passou a utilizar o MARC, surgiram variações de procedimentos que levaram a uma multiplicação de descrições em relação ao formato original. Surgiu então o formato UNIMARC para o intercâmbio internacional de registros MARC. As instituições nacionais produtoras de registros MARC adotavam normas locais, em seus países, e reformataavam seus registros em conformidade com o UNIMARC, para intercâmbio internacional. Recentemente, porém, vários fornecedores importantes de registros MARC concordaram em empregar o formato USMARC, de modo que é provável que, muito rapidamente, este venha a se tornar o padrão internacional. Portanto, o texto que se segue examina o formato USMARC.

O formato MARC inclui até 61 elementos de dados, 25 dos quais são recuperáveis diretamente durante as buscas. O formato é compatível com a última edição das Regras de Catalogação Anglo-Americanas e a edição mais recente da Classificação Decimal de Dewey, sendo previsível que venha a sofrer modificações, a fim de se adaptar a novas edições destes instrumentos de trabalho.

O formato MARC compreende duas seções: a seção 1, que contém informações descritivas dos dados bibliográficos; e a seção 2, que contém os dados bibliográficos propriamente ditos:

- título do registro
- diretório

- campos de controle
- campos de dados de tamanho variável.

Os campos abrangidos pela seção 2 e que, portanto, contêm os dados bibliográficos, são todos de tamanho variável. Por isso, é preciso sinalizar o começo e o fim de cada campo. Assim, cada campo é precedido de um parágrafo [*tag*] de três caracteres e dois indicadores numéricos, e termina com um delimitador especial. Os parágrafos consistem em três algarismos, situados na faixa 000-945, e têm uma estrutura mnemônica, pois obedecem à mesma seqüência do registro catalográfico, e os parágrafos das entradas secundárias refletem os parágrafos dos cabeçalhos principais. Os campos de tamanho variável são agrupados em blocos de acordo com o primeiro caractere do parágrafo:

- 1xx Entradas principais
- 2xx Títulos e parágrafo do título (título, edição, imprensa)
- 3xx Descrição física, etc.
- 4xx Indicação de série
- 5xx Notas
- 6xx Campos de acesso de assuntos
- 7xx Entradas secundárias exceto assunto ou série
- 8xx Entradas secundárias de série
- 9xx Dados de local.

Criam-se parágrafos para campos específicos pela inserção de dígitos nas duas últimas posições. Assim, por exemplo, adotam-se os seguintes parágrafos:

- 100 Entrada principal de autor pessoal
- 110 Entrada principal de nome-entidade [autor coletivo]
- 240 Título uniforme
- 245 Título e indicação de responsabilidade
- 250 Edição e indicação de autor, organizador, etc. da edição
- 260 Imprensa.

Além disso, o nome de um autor pessoal em geral leva '00' na segunda e terceira posições, de modo que, por exemplo,

- 100 é usado para uma entrada principal de autor pessoal
- 600 é usado para uma entrada de autor pessoal como cabeçalho de assunto
- 700 é usado para uma entrada secundária de autor pessoal.

Cada um dos campos principais contém também dois indicadores de campo, que são algarismos de um dígito colocados em seguida ao parágrafo e excluídos do campo ao qual são atribuídos. Empregam-se os indicadores para distinguir tipos diferentes de informações inseridas no mesmo campo, acolher entra-

das secundárias de título, indicar o número de caracteres a serem desprezados na alfabetação dos títulos, e mostrar se certas informações, como edição e imprensa, se relacionam com uma parte ou com o todo de obra em múltiplas partes. Por exemplo, no campo destinado ao cabeçalho de entrada principal de autor-entidade, utilizam-se os seguintes indicadores junto com o parágrafo 110:

- 110.00 Entrada invertida de autor coletivo
- 110.10 Entrada de governo
- 110.20 Entrada em ordem direta de autor-entidade.

Muitos campos num registro catalográfico contêm unidades distintas menores, conhecidas como subcampos. Os usuais na área de publicação (imprensa) são: lugar de publicação, publicador (editora) e data de publicação. Todos os campos são precedidos de um código de subcampo, que consiste num único símbolo (por exemplo, '\$') e uma única letra. Uma área de publicação seria codificada assim: '260.00 \$a São Paulo \$b Melhoramentos \$c 1993'.

Os códigos de subcampo são definidos no contexto do campo em que são empregados, mas códigos similares são usados em situações paralelas. Por exemplo, os códigos de subcampo para um nome pessoal são constantes, independentemente de ele ou ela ser autor ou assunto, principal ou secundário.

Concluída esta explicação sobre campos de dados de tamanho variável, voltemos à seção 1. Os campos de controle são a única parte da seção 1 que é inserida pelo catalogador. Contêm dados como: número de controle do registro (por exemplo, ISBN), língua do texto, código de nível intelectual, código do país de publicação, e o controle do acesso ao registro principal.

Cada registro começa com um rótulo e um diretório, ambos fornecidos pelo programa. O rótulo contém informações sobre o registro, como, por exemplo, seu tamanho e situação (novo, modificado, etc.), tipo e classe. O diretório é uma lista de localização que relaciona, para cada parágrafo, o parágrafo, o número de caracteres do campo e a posição do caractere inicial dentro do registro.

A estrutura do registro MARC é deliberadamente complexa, para permitir o máximo de flexibilidade. Quase todo elemento pode ser usado como ponto de acesso, e cada elemento ter qualquer tamanho. Tal complexidade foi prevista numa época em que a norma eram os catálogos impressos ou em microfime. Esse formato deu enorme contribuição à padronização e à comunicação por redes, mas há quem creia que chegou a hora de reavaliá-lo. Talvez um formato diferente seja mais apropriado num ambiente onde predominem os catálogos em linha de acesso público. A figura 5.2 (p. 108) mostra um registro MARC.

**Reflexão** Qual a função de cada um dos itens de codificação do formato de registro MARC?

### O Formato Comum de Comunicação

Existem muitos formatos de registros bibliográficos. As descrições bibliográficas adotadas por esses formatos diferem grandemente entre si, dependendo de sua origem. Os serviços de indexação e resumos empregam regras de descrição bibliográfica diferentes das adotadas na catalogação feita em bibliotecas. O formato MARC, que é utilizado como formato de intercâmbio por importantes bibliotecas, pressupõe como norma a Descrição Bibliográfica Internacional Normalizada [*International Standard Bibliographic Description* (ISBD)]. Por outro lado, os serviços de indexação e resumos (mas não todos) adotam o manual de referência UNISIST, que prescreve seus próprios designadores de conteúdo para as descrições bibliográficas de vários tipos de materiais. Esses dois importantes formatos definem, organizam e identificam os elementos de dados de modo diferente e se baseiam em diferentes conjuntos de códigos. Dessa forma, tem sido difícil ou praticamente impossível reunir num único arquivo registros bibliográficos procedentes de origens distintas. O Formato Comum de Comunicação [*Common Communication Format* (CCF)] foi projetado, então, com o objetivo de facilitar a comunicação de dados bibliográficos entre setores da área de informação. Em comum com o MARC, o CCF constitui uma implementação específica da norma ISO 2709. Assim, ele:

- especifica um pequeno número de elementos de dados obrigatórios, aceitos como essenciais para a identificação de um item
- proporciona elementos obrigatórios que são suficientemente flexíveis para comportar diferentes maneiras adotadas na descrição bibliográfica
- proporciona vários elementos optativos
- permite à instituição onde se origina o registro incluir elementos não-padrionizados mas que sejam considerados úteis em seu sistema
- proporciona um mecanismo para ligação de registros e segmentos de registros sem impor à instituição que lhes dá origem um método uniforme relativo ao tratamento de grupos de registros ou elementos de dados afins.

No Formato Comum de Comunicação, cada registro compreende quatro partes:

- rótulo de registro
- diretório
- campos de dados
- separador de registros.

### Outros formatos de registro bibliográfico em aplicações não-catalográficas

Além das aplicações em que se adotam formatos padronizados, existem várias outras que lançam mão de formatos não-padrionizados. Fundamentalmente,

encontram-se duas categorias diferentes de sistemas: bases de dados que se acham disponíveis para o público e sistemas de uso localizado suportados por pacotes de programas.

No caso das grandes bases de dados acessíveis ao público, são poucas as pressões no sentido de que seja aceito um formato-padrão, e cada produtor, geralmente, escolhe um formato de registro que seja adequado a essas bases de dados. Até mesmo pode ocorrer que uma mesma base de dados se apresente com diferentes formatos de registro conforme o serviço de buscas em linha que a torne disponível. As decisões em cada caso são tomadas tendo em vista os campos a serem incluídos e a indexação temática oferecida. Contudo, outra variável que se coloca é a presença de bases de dados de texto integral, e, mais recentemente, de multimídia, que pedem um formato de registro um tanto diferente dos registros bibliográficos para que as informações sejam apresentadas de forma adequada.

Os formatos de registro que se encontram em sistemas locais suportados por pacotes de programas são inúmeros e variados. A maioria desses pacotes de programas oferece sistemas de catalogação que aceitarão um formato de registro MARC ou que produzirão registros que sejam compatíveis com o mesmo formato MARC. Já outros não oferecem essa opção. Praticamente todos os pacotes de programas oferecem a seus compradores a oportunidade de desenvolver um formato de registro que seja adequado a uma aplicação específica. Assim, é possível haver em sistemas locais uma grande variabilidade de formatos de registro, à medida que se implementam projetos de acordo com os parâmetros estabelecidos pelos vários pacotes de programas.

### Metadados

Metadados quer dizer dados relativos a dados. Os registros bibliográficos são um tipo de metadados. Cada vez mais, porém, eles vêm sendo empregados no contexto mais especializado de dados que se referem a recursos digitais disponíveis numa rede. Os metadados também diferem dos dados bibliográficos ou catalográficos porque a informação de localização acha-se contida no registro de modo a possibilitar o fornecimento direto do documento a partir de programas aplicativos apropriados, ou, em outras palavras, os registros podem conter informações detalhadas quanto ao acesso e endereços da rede. Além disso, os registros bibliográficos são projetados para que os consulentes os utilizem não apenas quando estiverem avaliando a relevância e decidindo se desejam localizar o recurso original, mas também como um identificador único do recurso, de modo que possam solicitá-lo — recurso ou documento — numa forma que tenha sentido para a pessoa a quem a solicitação for entregue. Essas funções continuam sendo importantes. Os mecanismos de busca da Internet (ver capítulo 8) empregam metadados nos processos de indexação que adotam para indexar recursos da Internet. Os metadados precisam ter condições de

descrever localizações e versões de documentos que estejam em pontos remotos, bem como adaptar-se à falta de estabilidade da Internet, dados redundantes, perspectivas diferentes quanto à granularidade da Internet (o que é um documento, ou um recurso?), e localizações diversas em várias redes diferentes.

### 5.5 Estruturas das bases de dados

A forma mais rudimentar de busca numa base de dados é percorrê-la, registro por registro, à procura do elemento de dado apropriado. No caso de grandes arquivos, na etapa atual dos conhecimentos, isso seria moroso, pelo que foram desenvolvidos métodos alternativos para a localização de registros específicos.

Os serviços de buscas em linha e outras aplicações que empregam sistemas de gerenciamento de documentos sempre utilizaram o método de arquivo invertido, descrito mais adiante, e que é útil nas buscas em bases de dados textuais complexas, pois quem faz as buscas ignora a forma como a chave de busca foi inserida na base de dados e, fundamentalmente, precisa adivinhar aquela que parece ser a mais apropriada.

Os sistemas de processamento de transações, como os de gerenciamento de bibliotecas, reservas de passagens e informação de vendas e *marketing* podem também empregar esse método para localizar registros específicos numa base de dados, os quais, porém, também precisam de um mecanismo que faça a ligação entre uma série de distintas bases de dados, de modo que as informações sejam extraídas de mais de uma base para serem exibidas no monitor ou impressas. Este requisito imposto aos sistemas de processamento de transações acarretou o desenvolvimento de estratégias de otimização do projeto de bases de dados. A seguir descrevemos alguns desses métodos.

#### Listas enlaçadas unidirecionais

As listas enlaçadas unidirecionais são uma forma simples de localizar um registro específico. Tratamos delas aqui com o objetivo de oferecer uma base para o exame, mais adiante, dos arquivos invertidos. As listas enlaçadas unidirecionais usam um campo-chave, por exemplo, o campo de autor, para localizar um registro específico. Os registros podem ser dispostos em ordem pelo campo-chave, embora isso seja inconveniente numa área sujeita a mudanças. O mais comum é os registros serem armazenados numa ordem diferente, empregando-se um sistema de apontadores para representar o arranjo. O processamento é feito seguindo-se esses apontadores. A figura 5.5 contém um diagrama disso.

Este método baseia-se numa única chave e não será satisfatório se a busca for realizada em outro elemento de registro. Por exemplo, quando o campo-chave é o de autor, os apontadores não suportarão buscas nos campos de título ou assuntos. Exige-se que muitos sistemas suportem buscas em mais de uma chave, pelo que foram desenvolvidos sistemas multiindexados que empregam

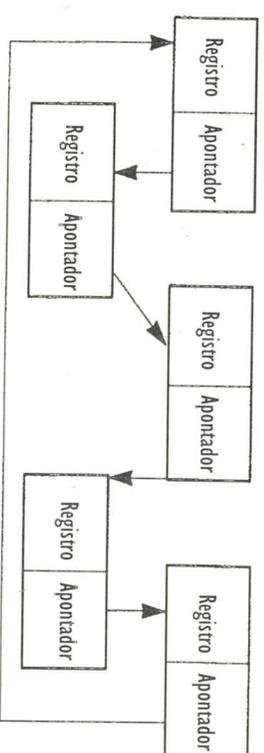


Figura 5.5 Uma lista enlaçada unidirecional

vários índices para apontar registros no campo principal. Muitos de tais sistemas empregam arquivos invertidos.

#### Arquivos invertidos

O arquivo invertido é similar a um índice. No método de arquivo invertido pode haver dois ou três arquivos separados. Aqui, tendo em vista a apreensão do conceito básico, descreveremos o método de dois arquivos, que emprega o arquivo de texto ou impresso e o arquivo inverso ou de índice.

O **arquivo de texto** contém os registros propriamente ditos. O **arquivo de índice** proporciona acesso a esses registros. O arquivo de índice contém um registro de cada um dos termos indexados de todos os registros constantes da base de dados, dispostos em ordem alfabética. Cada termo é acompanhado de informações sobre sua frequência de ocorrência na base de dados, o arquivo onde se acha localizado, o registro no qual está inserido, e possivelmente outras informações quanto à localização, como o parágrafo (ou campo) no qual se acha presente. Ao acrescentar um novo registro à base de dados, é preciso atualizar o arquivo de índice.

Esses arquivos são empregados de modo conjunto durante as buscas na base de dados. Se o usuário estiver interessado em fazer uma busca, por exemplo, sobre a palavra 'muros', digitará o termo no teclado e o sistema procurará esse termo no arquivo de índice. Se o termo não estiver presente no arquivo de índice, o sistema responderá indicando que não existem ocorrências desse termo. Se for encontrado, o usuário será informado sobre quantas ocorrências, ou lançamentos, do termo existem na base de dados. Para que os registros sejam exibidos, utiliza-se a localização no arquivo de texto para localizar registros nesse arquivo.

No caso da utilização de três arquivos, existe um arquivo intermediário que permite que a digitação dos termos de busca no teclado seja conferida rapidamente, e apresentado na tela o número de ocorrências. Isto é particularmente útil no caso de buscas complexas que envolvam o uso dos registros do arquivo de índice para vários termos de busca.

O objetivo da descrição acima é oferecer uma introdução simples ao conceito básico de arquivo invertido. Na prática, as estruturas de arquivos podem ser mais complexas. Por exemplo, para que seja possível fazer buscas de termos que guardem proximidade com outros termos (por exemplo, termos separados por duas outras palavras), o arquivo de índice deverá conter informação sobre as posições das palavras dentro de um campo para cada termo.

Muitas vezes os arquivos invertidos são criados para vários campos dentro de um registro. No entanto, nem todos os campos são em geral indexados, pois cada índice ocupa espaço de memória no disco; os índices são criados para aqueles campos em que regularmente se realizam buscas. Usualmente criam-se arquivos invertidos para nomes de autores, palavras dos títulos, termos de indexação de assuntos e formas abreviadas de autor e título.

A estratégia de recuperação acima descrita depende de sua possibilidade de recuperar registros específicos e subseqüentemente identificar campos dentro desses registros. Extensos registros de texto integral precisam ser subdivididos em parágrafos e a estes devem ser atribuídos identificadores, antes de se começar a indexar. Alternativamente, as posições de palavras específicas no arquivo podem ser usadas como identificadores.

Algumas bases empregam arquivos de autoridade, na forma de listas de entradas de autor ou cabeçalhos de assuntos, ou tesouros. Neste caso, haverá uma ligação entre o arquivo de índice ou o arquivo de termos e a lista de autoridade, de modo que sejam exibidos os termos autorizados e as relações entre eles.

### **Reflexão** São necessários arquivos de autoridade para criação de índices de documentos textuais?

### **O desenvolvimento de estruturas de bases de dados**

Nos primórdios da informática, os sistemas destinados a empresas e bibliotecas funcionavam com uma série de arquivos-mestre específicos, abrangendo, por exemplo, no caso de bibliotecas, leitores registrados e livros do acervo, ou, no caso de empresas, folha de pagamento, vendas e inventário. Quando as empresas utilizavam computadores em diferentes aplicações e nem todas as atividades que poderiam ter sido estavam informatizadas, isso funcionava satisfatoriamente. No entanto, logo ficou evidente que programas destinados, por exemplo, ao controle de circulação numa biblioteca precisavam acessar dois arquivos diferentes, ou até mais, e que seria apropriado começar a examinar as relações entre esses arquivos. Isso levou à introdução do conceito de base de dados, junto com o programa destinado a gerenciá-la, conhecido como sistema de gerenciamento de bases de dados. Tornou-se então necessário estudar a melhor maneira de estruturar os dados ou desenvolver modelos de dados que suportassem aplicações específicas. De modo mais formal:

- 1 **Uma base de dados** é uma coleção geral e integrada de dados junto com descrição deles, gerenciada de forma a atender a diferentes necessidades de seus usuários.
- 2 **Um sistema de gerenciamento de bases de dados (SGBD)** é o sistema que gera, opera e mantém bases de dados, e, como tal, deve incluir todos os programas necessários para essa finalidade.
- 3 **Um modelo de dados** especifica as regras segundo as quais os dados são estruturados, bem como as operações correlatas que são permitidas. Pode ser também visto como uma técnica para a descrição formal de dados; relações entre dados e limitações de uso.

### **Características das bases de dados**

Uma das finalidades principais de uma base de dados é que os dados que ela contém sirvam a uma variedade de aplicações distintas. Para que isso seja possível, é importante que a base de dados apresente as seguintes características

- 1 Ser substancialmente não-redundante, isto é, possuir o mínimo de duplicidade de dados idênticos, de preferência nenhuma. Essa duplicidade acarreta dificuldade para se manter a consistência de dados, principalmente durante a atualização, e o desperdício de espaço de armazenamento.
- 2 Ser independente de determinado programa (conceito conhecido como independência de dados), de modo que os dados possam ser transferidos ou reestruturados sem a necessidade de fazer alterações nos programas.
- 3 Ser utilizável por todos os programas.
- 4 Incluir todas as inter-relações de dados que forem necessárias, de modo e suportar a variedade de usos que podem ser atribuídos aos dados.
- 5 Possuir um método comum de recuperação, inserção e correção de dados.

### **Reflexão** Por que existe a possibilidade de a duplicidade de dados causar inconsistências nos registros de dados? (Pense no processo de atualização.)

### **Dados lógicos e físicos**

No método de base de dados, embora a base propriamente dita possa estar armazenada fisicamente como um conjunto de arquivos, os usuários e os aplicativos não precisam saber sobre o armazenamento físico dos dados. Armazenada com os dados encontra-se uma descrição da base de dados que permite ao SGBD recuperar informações e armazenar novos dados em lugares apropriados da base, estabelecendo relações com outros dados, se forem pertinentes. Os aplicativos não acessam diretamente a base de dados; ao contrário, transferem as consultas ao SGBD para que recupere ou armazene os dados. Em essência, o SGBD é orientado aos dados. Essa mudança para a orientação aos dados está...

muito maior importância à modelagem de dados e ao projeto de base de dados, ficando o armazenamento físico dos dados por conta do SGBD e não do projetista de aplicativos do sistema.

### Reflexão **Explique a diferença entre estruturas físicas e lógicas de bases de dados.**

#### Estruturas lógicas de bases de dados

Quando é preciso acessar uma série de arquivos interligados, como num sistema de gerenciamento de bibliotecas, deve-se dispor de diretrizes para alocar dados a arquivos específicos no sistema da base de dados e definir os melhores vínculos ou chaves entre os arquivos. São três os principais tipos estruturais de bases de dados e SGBD correlato: hierárquicas, em rede e relacionais.

#### Bases de dados hierárquicas

As bases de dados hierárquicas são estruturadas como um arranjo arborescente. Cada item de dados localiza-se num item de dados de nível superior. Assim, se a base de dados contiver informações sobre os empregados da empresa, elas serão organizadas por fábricas, departamentos e em seguida os empregados dos diferentes escalões. Tal método permite a rápida recuperação dos dados, desde que se conheçam as informações de nível superior, como o departamento onde o empregado trabalha. Se essas informações não forem conhecidas, será difícil fazer buscas numa estrutura hierárquica. Para todo livro catalogado haverá informações relativas a cada exemplar e cada reserva no primeiro nível da hierarquia. Para cada exemplar do livro, haverá informações sobre empréstimo bem como sobre a pessoa a quem o livro foi emprestado, que incluirão dados sobre a situação do empréstimo. Para cada reserva haverá informações sobre em nome de quem foi feita a reserva, e, novamente, dados sobre a situação do empréstimo. Nesse modelo os dados serão acessados por intermédio de 'livro catalogado'. Para encontrar dados sobre determinada pessoa, por exemplo, será preciso vasculhar a base inteira, pois não haverá acesso direto pela pessoa.

#### Bases de dados em rede

O método de rede baseia-se em vínculos explícitos ou apontadores entre entidades afins. O modelo de rede mais conhecido é o Codasy1. Esta abreviatura significa Committee on Data Systems Languages, que é o órgão responsável pela concepção desse modelo. No modelo em rede existe entre os itens de dados nos vários níveis um vínculo mais direto do que no modelo hierárquico, por meio do emprego de apontadores que ligam os dados em níveis diferentes. O método de rede requer um grande número de vínculos entre os itens de dados, que ocupam uma parcela considerável de espaço de armazenamento.

#### Bases de dados relacionais

As bases de dados relacionais empregam um tipo de estrutura de base de dados que tem sido amplamente utilizado em sistemas de bases de dados. Nos sistemas relacionais, as informações são mantidas num conjunto de relações ou tabelas. As fileiras das tabelas equivalem a registros, e as colunas, a campos. Os dados nas várias relações são interligados mediante uma série de chaves. A figura 5.7 mostra um exemplo simples de uma relação conhecida como catalogação-livro. Nesta relação o ISBN é a chave primária e pode ser usado em outras relações para identificar um livro específico. Por exemplo, se mantivermos a relação encomenda-livro, o ISBN funcionará como um vínculo com a relação encomenda-livro. Se desejarmos preencher um formulário de encomenda com informações do arquivo de encomendas, será possível extrair do arquivo de catálogo dados relativos a cada livro, que serão impressos nos formulários de encomenda junto com dados oriundos do arquivo de encomendas.

As bases de dados relacionais são construídas com o emprego de uma técnica de análise de dados chamada 'normalização'. A normalização é usada para decompor os dados em tabelas de modo que os campos de cada tabela sejam dependentes apenas do campo-chave e não se vinculem a qualquer outra chave. As relações são representadas por duplicação dos itens de dados. A normalização permite que sejam feitas inserções, eliminações e correções de dados sem consequências indesejáveis.

Ocorrências da relação catalogação-livro			
ISBN	Título	Autor	Ano
0-82112-462-3	Organic chemistry	A.J. Brown	1989
0-84131-460-7	Alchemy	R.M. Major	1987
0-69213-517-8	Expert systems	S. Estelle	1988
0-93112-462-1	Computer science	S. Estelle	1989
0-71143-526-6	Bibliography	J. Johns	1991

Ocorrências da relação encomenda-livro		
Pedido nº	ISBN	Quantidade encomendada
678	0-82112-462-3	1
678	0-84131-460-7	4
678	0-69213-517-8	20
679	0-93112-462-1	2
680	0-82112-462-3	3
681	0-71143-526-6	2

Figura 5.7 Duas relações simples

## Estruturas de bases de dados multimídia

As bases de dados multimídia colocam novos desafios no que tange à sua estrutura. Como fotografias, animação, som, texto e tabelas de dados possuem necessidades de armazenamento muito diferentes, os sistemas de gerenciamento de bases multimídia procuram empregar uma gama de tecnologias, como a tecnologia relacional para tabelas, bases de dados textuais para documentos, e dispositivos de armazenamento de imagem para gráficos e animação. Um problema crucial é processar itens não-textuais, como desenhos e imagens em movimento. Em mídia de tempo variável, os padrões de vídeo digital interativo proporcionam acesso por quadros. Para acesso mais elaborado, as imagens têm de ser indexadas com palavras-chave, como se fossem documentos textuais.

## Estruturas de bases de dados orientadas a objetos

Os sistemas de gerenciamento de bases de dados orientadas a objetos (SGBD-OO) oferecem recursos para armazenamento de itens criados por programas orientados a objetos. Uma vez que objeto é uma coleção complexa de itens de dados, que se definem como estando relacionados com outros objetos, os SGBD-OO precisam oferecer recursos para grandes objetos que são organizados em hierarquias segundo suas relações com outros objetos. Essas estruturas possuem algo em comum com o modelo hierárquico. A tecnologia SGBD-OO ainda está em desenvolvimento, sendo improvável que venha a tomar o lugar da tecnologia relacional nas aplicações comerciais tradicionais, pois não oferece as mesmas vantagens de gerenciamento de dados. No entanto, quando os itens de dados são complexos e heterogêneos, como ocorre nas bases de dados multimídia, talvez os SGBD-OO sejam uma possibilidade interessante.

## RESUMO

Os dados no computador são armazenados em arquivos, que abrangem uma série de registros subdivisíveis em campos. Chama-se base de dados um grupo desses arquivos ou determinado arquivo que suporte aplicação específica. Há vários tipos de bases de dados, inclusive as bibliográficas e de fontes. As bases de dados de referências incluem as de dados bibliográficos e de dados catalográficos. A estrutura do registro influi na forma como os dados serão posteriormente recuperados. Os formatos MARC e CCF são formatos de registro especiais encontrados em bases de dados bibliográficos. Há vários formatos de registro em outras aplicações. Para recuperar os dados é preciso que as bases sejam estruturadas e indexadas adequadamente. Os arquivos invertidos são bastante adotados em sistemas de gerenciamento de bases de dados, que adotam um de três métodos na estruturação lógica das bases: hierárquico, em rede e relacional. As bases multimídia e orientadas a objetos são importantes em aplicações especiais.

## QUESTOS DE REVISÃO

- 1 Explique a diferença entre estruturas de registro baseadas em campos de tamanho variável e as de campos de tamanho fixo.
- 2 Relacione os diferentes tipos de bases de dados bibliográficos.
- 3 Por que o formato MARC é importante?
- 4 Descreva sucintamente:
  - uma lista enlaçada unidirecional
  - um arquivo invertido.
- 5 Quais as principais estruturas lógicas de bases de dados? Quais as vantagens e desvantagens de cada uma?

## Bibliografia

### Formato de registro MARC

- Anderson, D. *Standard practices in the preparation of bibliographic records*. London IFLA UBCIM, 1989. (UBCIM Occasional Papers 13)
- BEDIS. *Report of the Working Party of the UK Book Trade Electronic Data Interchange Standards Committee*. London: Whitaker, 1990.
- Bierbaum, E. G. MARC in museums: applicability of the revised visual materials format *Information technology and libraries*. v. 9, n. 4 p. 291-299, Dec. 1990.
- Bourne, R. UKMARC—a format for the 21st century. *New library world*, v. 93, n. 1101 p. 4-8, 1992.
- Gredley, E.; Hopkinson, A. *Exchanging bibliographic data: MARC and other international formats*. Ottawa: Canadian Library Association; London: Library Association 1990.
- Holt, B. P. et al. (ed.) *UNIMARC manual*. London: IFLA UBCIM, 1987.
- McCallum, S. H.; Roberts, W. D. (ed.) *UNIMARC in theory and practice: papers from the UNIMARC Workshop, Sydney, August 1988*. London: IFLA, 1989.
- Plessard, M.-F. The Universal Bibliographic Control and International MARC Programme. *International cataloguing and bibliographic control*, p. 35-37, 1990.
- Sweeney, R. Survey of the use of UKMARC 1991: selected results. *New library world*, v. 92, n. 1097, p. 5-9, 1991.

### Outros formatos de registro e estruturas de bases de dados

- Ashford, J. A.; Willet, P. *Text retrieval and document databases*. Bromley: Chartwell Bratt, 1989.
- Dierckx, H. The UNISIST reference manuals and UNIBID standardisation for development. *Program*, v. 17, n. 2, p. 68-85, 1983.
- ; Hopkinson, A. *Reference manual for machine-readable bibliographic descriptions*. Paris: Unesco/General Information Programme and UNISIST, 1981.
- Flynn, R. *An introduction to information science*. New York: Dekker, 1987.
- Kemp, A. *Knowledge based systems*. London: Ashb, 1990.
- Oxborrow, E. A. *Databases and database systems: concepts and issues*. Bromley: Chartwell Bratt, 1991.