

ESTRUTURA DE BASES DE DADOS: MODELOS DE METADADOS E A QUALIDADE DE RESPOSTA

DATA BASES' STRUCTURE: METADATA MODELS AND ANSWER QUALITY

Marta Lígia Pomim VALENTIM*

RESUMO

A estrutura das bases de dados, bem como a forma com que as bases de dados desenham seu modelo de dados são objetos de análise deste artigo. Alguns tipos de bases de dados, como: bibliográficas, diretórios, financeiras, numéricas e texto completo, são analisadas a partir das diferentes tipologias documentais que as compõem e a relação com o desenho do modelo de dados utilizados por elas. A partir disso, defende-se a hipótese de que quanto melhor desenhado o modelo de dados em relação as diferentes tipologias documentais, maior será a qualidade de resposta da base de dados.

Palavras-chave: Estrutura de Bases de Dados; Modelo de Dados; Metadados; Tipologias Documentais; Qualidade de Resposta.

ABSTRACTS

The structure of databases, as well as the way that databases design their data model are objects of analysis in this article. Some types of databases like: bibliographical, directories, financial, numeric and full text are analyzed according to the different document typologies that compose them and the relation between their data model design. On these grounds, it defends the hypothesis that the better the data model designed in relation with the different documental typologies, the higher the answer quality of database will be.

Key-words: DataBase' Structure; Data Model; Metadata; Documental typologies; Answer Quality.

Na década de sessenta, as instituições que processavam a informação, selecionando-a, analisando-a e disponibilizando-a de forma impressa, como os conhecidos *abstracts* ou *index*, perceberam a

Na década de sessenta, as instituições que processavam a informação, selecionando-a, analisando-a e disponibilizando-a de forma impressa, como os conhecidos *abstracts* ou *index*, perceberam a

(*) Doutora em Ciência da Informação e Documentação pela ECA/USP. Autora do livro "O Custo da Informação Tecnológica". Organizadora e uma das autoras do livro "O Profissional da Informação: formação, perfil e atuação profissional". Líder do Grupo de Pesquisa "Interfaces: Informação e Conhecimento". Docente da Universidade Estadual de Londrina. Presidente da Associação Brasileira de Ensino em Biblioteconomia, Documentação e Ciência da Informação (ABEBD), gestão 2001 - 2004. Atuou como Coordenadora de Informação Tecnológica na indústria Metal Leve S/A durante oito anos.

Endereço: Rua Euclides da Cunha, 491 - Bairro Shangri-lá A CEP 86070-500 - Londrina - PR - Tel.: (0xx43) 327-2684 - E-mail: valentim@uel.br

necessidade de utilizar o recurso da informática para agilizar o acesso às informações.

Essas instituições aliaram-se ao desenvolvimento da informática - na época em que se iniciava a democratização do uso dos microcomputadores nos Estados Unidos e Europa - década de sessenta - e passaram a processar a informação de forma automatizada. A esse novo formato de organização e disponibilização da informação dá-se o nome de bases de dados.

Observa-se, no desenvolvimento histórico apresentado na Figura 1, que as bases de dados surgem primeiramente em suportes impressos, designados de: *abstracts*, *index*, bibliografias, diretórios, anuários, guias, cadastros, enfim vários tipos de publicações que congregavam informação de um mesmo tipo documental, para serem disseminadas a uma determinada clientela. A atualização era difícil e o custo muito alto, por causa dos custos do papel e da impressão; existia também a falta de padronização de formatos; além disso, a disseminação era restrita a públicos considerados “de elite”, ou seja, comunidades científicas ou industriais que tinham recursos para acessar a informação.

O surgimento, na década de sessenta, dos sistemas de informações *on-line* que disponibilizavam a informação em meio magnético, através de redes cooperativas, mais tarde, aliados à telecomunicação, tornam-se sistemas de informações eletrônicos. A automação resolve, em parte, muitos dos problemas citados anteriormente, inclusive a questão da padronização de formatos. Nessa época, com o avanço das telecomunicações e das tecnologias da informação, inicia-se a democratização do acesso à informação,

mas é importante frisar que essa democratização se dá comparativamente ao período anterior, uma vez que até os dias de hoje ainda é uma pequena parcela da população que acessa esse tipo de sistema de informação.

Para os países desenvolvidos que possuíam e dominavam a telemática (informática + telecomunicações), o acesso era rápido, os custos eram acessíveis para qualquer instituição e, até mesmo, para pessoas físicas. Todavia, para os países em desenvolvimento e para os países subdesenvolvidos, o acesso era lento, porquanto não possuíam telemática de qualidade e os custos de acesso tornavam-se muito altos.

A partir dessa constatação, os sistemas de informação comerciais como o Dialog, por exemplo, perceberam a necessidade de disponibilizar bases de dados, utilizando novas tecnologias que não necessitassem da telecomunicação, surgindo dessa maneira a tecnologia óptica. Essa nova tecnologia atendia uma demanda de países que precisavam acessar a informação e que tinham no custo da telecomunicação, um fator inibidor na democratização desse acesso.

A Internet revoluciona o acesso aos sistemas de informação, além de redimensionar a linguagem utilizada, a forma de apresentação e a interatividade das multimídias disponíveis na rede. Apesar de a rede Internet ainda ser acessada por uma minoria populacional, ela democratizou o acesso de pessoas e organizações ao mundo eletrônico/digital informacional. Observa-se uma tendência de ampliação do acesso, mesmo que de forma lenta, uma vez que as tecnologias da informação têm seu custo reduzido ano após ano.

Figura 1. Trajetória e Desenvolvimento das Bases de Dados

Tempo	Produtos	Tecnologia	Problemas
Anos 50	<ul style="list-style-type: none"> • Cadastros • Abstracts • Index • Diretórios • Anuários • Bibliografias • Balanços • Pesquisas • Etc. 	Os produtos são desenvolvidos e impressos em papel	<ul style="list-style-type: none"> • Falta de normatização e padrões; • Pouca cooperação entre instituições; • Difícil distribuição e disseminação da informação; • Alto custo de impressão e distribuição; • Demora no acesso à informação; • Qualidade na pesquisa questionável

Anos 60	<ul style="list-style-type: none"> • Cadastros • Abstracts • Index • Diretórios • Anuários • Bibliografias • Balanços • Pesquisas • Etc. 	<ul style="list-style-type: none"> • Os produtos são desenvolvidos e impressos em papel; • A cooperação acontece através da troca de fitas magnéticas 	<ul style="list-style-type: none"> • Início de normatização e definição de padrões; • Início de cooperação entre instituições; • Melhoria na distribuição e disseminação da informação; • Alto custo de impressão e distribuição • Maior velocidade no acesso à informação; • Qualidade na pesquisa questionável
Anos 70	Bases de dados eletrônicas do tipo: <ul style="list-style-type: none"> • Bibliográficas • Referenciais • Cadastrais • Numéricas 	<ul style="list-style-type: none"> • As bases de dados tradicionais mantêm o formato em papel, mesmo oferecendo o produto em formato eletrônico; • Uma parcela das novas bases de dados são oferecidas apenas em formato eletrônico, não tendo versão em papel; • A cooperação acontece através da troca de informações <i>on-line</i>, ou seja, utilização da microinformática e das telecomunicações 	<ul style="list-style-type: none"> • Normatização e padrões internacionalmente aceitos; • Cooperação entre instituições totalmente estabelecida; • Distribuição e disseminação da informação mais eficiente; • Alto custo de impressão e acesso às informações (telecomunicação); • Acesso em tempo real à informação; • Qualidade na pesquisa aceitável
Anos 80	Bases de dados eletrônicas e ópticas do tipo: <ul style="list-style-type: none"> • Bibliográficas • Referenciais • Cadastrais • Numéricas 	<ul style="list-style-type: none"> • As bases de dados tradicionais mantêm o formato em papel, mesmo oferecendo o produto em formato eletrônico e/ou óptico; • Uma parcela das novas bases de dados são oferecidas em formato eletrônico e/ou em formato óptico não tendo versão em papel; • A cooperação acontece através da troca de informações <i>on-line</i>, ou seja, utilização da microinformática, tecnologia óptica e das telecomunicações 	<ul style="list-style-type: none"> • Normatização e padrões internacionalmente aceitos; • Cooperação entre instituições totalmente estabelecida; • Distribuição e disseminação da informação mais eficiente; • Alto custo de impressão e acesso às informações (telecomunicação); • Baixo custo de acesso às bases de dados em formato óptico; • Acesso em tempo real à informação; • Qualidade na pesquisa aceitável
Anos 90	Bases de dados eletrônicas e ópticas do tipo: <ul style="list-style-type: none"> • Bibliográficas • Referenciais • Fontes (texto completo) <ul style="list-style-type: none"> - Numéricas - Textuais - Textuais e Numéricas 	<ul style="list-style-type: none"> • As bases de dados tradicionais mantêm o formato em papel, mesmo oferecendo o produto via Internet ou com tecnologia óptica; • Uma parcela das novas bases de dados são oferecidas em formato eletrônico via Internet e/ou em formato óptico não tendo versão em papel; • A cooperação acontece através da troca de informações <i>on-line</i>, ou seja, utilização da microinformática, tecnologia óptica e das telecomunicações, principalmente através do uso da rede Internet • Modelos de Metadados 	<ul style="list-style-type: none"> • Normatização e padrões internacionalmente aceitos; • Cooperação entre instituições totalmente estabelecida; • Distribuição e disseminação da informação muito mais eficiente; • Alto custo de impressão • Acesso às informações via Internet com custo subsidiado; • Baixo custo de acesso às bases de dados em formato óptico; • Acesso em tempo real à informação; • Qualidade na pesquisa aceitável; • Vários padrões de organização da informação no meio eletrônico/digital

A figura anterior demonstra como as bases de dados se desenvolveram: primeiramente surgem em formato papel, passam a ser oferecidas eletronicamente na década de setenta, na década seguinte passam a ser oferecidas também em formato óptico (*CD-ROM*) e,

finalmente, se reformulam eletronicamente para serem disponibilizadas pela rede Internet.

Para Takahashi (2000, p. 61) “a produção de conteúdos esbarra também em problemas como o alto custo da digitalização de acervos e as diferenças das

técnicas que envolvem a preparação de bases de dados a partir de formatos diversos”.

Rowley (1994, p. 66) define base de dados como “uma coleção de registros similares entre si e que contém determinadas relações entre esses registros”. Grossmann (1994, p. 95) define bases de dados como “qualquer coleção de informações agrupadas segundo um interesse comum e mantidas eletronicamente (em computadores)”. Continuando as definições de bases de dados, Teixeira e Schiel (1997, p. 66) afirmam que bases de dados são “fontes de informação automatizada que podem ser pesquisadas de diversos modos. Elas podem ser armazenadas em meios magnéticos ou ópticos e acessadas local ou remotamente”.

Lancaster (1993, p. 305) conceitua base de dados como:

“uma coleção de itens sobre os quais podem ser realizadas buscas com a finalidade de revelar aqueles que tratam de um determinado assunto. A base de dados consiste em artefatos, como livros (o acervo de uma biblioteca é uma base de dados com certeza), ou registros que representam os artefatos, como, por exemplo, registros bibliográficos constantes de páginas impressas, de fichas ou de meios eletrônicos”.

As bases de dados surgem justamente da necessidade de se obter informações com mais rapidez, mais qualidade, mais eficácia e de forma selecionada. A estrutura padronizada, congregando tipologias de bases de dados bem definidas, possibilita que sistemas de informações gerenciem uma grande quantidade de bases de dados de diferentes tipos e, ao mesmo tempo, disponibiliza ao usuário um serviço de qualidade quanto ao acesso e quantidade de informações.

O sistema de informação Dialog estabeleceu uma categorização para as 520 bases de dados que compõem o sistema atualmente, da seguinte forma: a) Bibliográficas; b) Diretórios; c) Financeiras; d) Numéricas; e) Texto Completo.

As bases de dados bibliográficas possuem uma estrutura bastante conhecida no meio científico; a informação é apresentada em formato de referência bibliográfica e, na maioria das vezes, contém o resumo dos documentos que a compõe.

A maioria das bases de dados bibliográficas trabalham com várias tipologias documentais como artigos de periódicos; *papers*; trabalhos científicos apresentados em congressos, simpósios, seminários etc.; patentes; normas técnicas; relatórios técnicos; catálogos entre outros. No entanto, algumas bases de dados bibliográficas trabalham com uma única tipologia

documental. As bases de dados bibliográficas são originárias das publicações impressas denominadas *abstracts*, *index* e bibliografias.

No caso de artigos de periódicos, considerados como tipologia documental, é importante lembrar que alguns títulos de periódicos oferecem não só a referência bibliográfica e o resumo, mas também o texto completo. Essa realidade ainda é incipiente em relação ao número de publicações periódicas existentes nas bases de dados; no entanto é uma tendência crescente. No sistema Dialog, atualmente, apenas 4 mil periódicos disponibilizam o texto completo de seus artigos científicos nas diferentes bases de dados que compõem o sistema.

Os critérios para uma publicação periódica ser aceita e indexada em uma base de dados bibliográfica varia de base para base, mas a maioria prima por selecionar publicações de qualidade para compor o rol de títulos indexados pela base de dados. Podem-se citar alguns critérios mais comuns, solicitados por esse tipo de base de dados às publicações científicas: a) periodicidade deve ser respeitada; b) comissão editorial ou corpo editorial para processar a avaliação e seleção dos artigos publicados pelo periódico; c) fabricação da informação documentária (palavras-chave e resumo) deve obrigatoriamente ser elaborada.

A estrutura das bases de dados bibliográficas tem em comum o modelo de metadados e a condensação da informação, ou seja, a representação e descrição dos dados que identificam o documento propriamente dito e a representação e descrição do conteúdo do documento através de descritores/palavras-chave e resumos: estes têm o objetivo de dar acesso à informação, recuperando-a de maneira eficaz.

Como exemplo de bases de dados bibliográficas, cita-se primeiramente a base de dados Medline, que conforme característica mencionada anteriormente, congrega diferentes tipologias documentais, dentre as quais: a) artigos de periódicos; b) relatórios governamentais; c) cartas; d) relatórios técnicos; e) anais de eventos; f) *papers*; g) monografias; h) publicações especiais; i) teses.

Portanto, verificam-se nove diferentes tipologias documentais que compõem a base, situação essa encontrada na maioria das bases de dados desse tipo, diferentemente da base de dados bibliográfica Dissertation Abstracts Online que congrega apenas uma única tipologia documental.

O registro informacional da base de dados Medline atende a nove diferentes tipologias documentais com o mesmo modelo de dados, isto é, as diferentes

tipologias documentais que compõem a base têm de se adequar, em termos de representação e descrição, ao modelo único utilizado pela base.

Essa é a realidade da maioria das bases de dados bibliográficas existentes e, de certa forma, questiona-se esse modelo de dados em relação à completude informacional obtida no resultado final da pesquisa, ou seja, se cada tipologia documental tem descrições específicas e obedece suas próprias características documentais, um único modelo de dados dificilmente atenderá a todas as especificidades de diferentes tipos documentais.

Outra questão relevante é em relação aos conteúdos temáticos específicos, encontrados de formas diferentes na estrutura textual de cada tipologia documental. A questão requer, portanto, uma adequação dos conteúdos temáticos ao modelo de dados oferecido pela base de dados.

O modelo de metadados das bases de dados bibliográficas deve, conseqüentemente, ser revisto, buscando maior completude informacional. Quanto mais completo o modelo de dados, maior será a completude de informações obtidas em uma pesquisa. Dessa forma, será possível oferecer uma recuperação com maior qualidade de resposta que satisfaça realmente às necessidades informacionais do usuário. Nesse sentido, vale dizer que o usuário tem sempre um objetivo próprio de pesquisa e, nem sempre, necessita da totalidade de informações. No entanto, caso ele necessite da totalidade de informações, elas estarão disponíveis para acesso e uso.

O registro informacional da base de dados Medline (<http://library.dialog.com/bluesheets/html/b10154.html>), compreende o seguinte modelo de metadados:

Figura 2. Modelo de Metadados - Base de Dados Medline

AN	Numero de acesso do Dialog
TI	Título do documento
AU	Autor/Autores do documento
CS	Identificação da fonte responsável pela informação
JN, CP, PY, SO, SN	JN=Título do periódico, CP= País da publicação, PY= Data da publicação, SO= Casa publicadora/ Editora, SN= ISSN
JC	Código do periódico
CN, CA	CN= Número de contrato, CA= Número de chamada
NT	Notas/Comentários
LA, SL	LA= Língua do documento, SL= Língua dos dados na base
DT	Tipo de documento
JA	Anúncio do periódico
SF	Subcampo
AB	Resumo
GS	Checagem da citação
DE	Descritores/Palavras-chave
SQ	Número do Banco de Seqüência Molecular
ID, RN	ID= Identificadores, RN= Número de registro da Chemical American Society (CAS)
ID, EC	ID= Identificadores, EC= Número da Comissão de Enzima
NM	Nome de pessoa

Constata-se que o modelo de metadados da base Medline é composto por 27 campos informacionais que atendem, conforme já mencionado anteriormente, nove diferentes tipologias documentais.

As bases de dados do tipo diretório possuem estrutura bastante diversificada, porque atendem a diferentes tipologias documentais e respeitam essas diferentes estruturas textuais. Dessa maneira, as especificidades de representação e descrição dos dados que identificam o documento propriamente dito e da representação e descrição do conteúdo do documento

são integralmente absorvidas e disponibilizadas para acesso e uso do público a que se destina quando da recuperação da informação.

As informações que compõem as bases de dados do tipo diretório variam desde dados referenciais sobre pessoas como, por exemplo, a base de dados Marquis Who's Who; instituições públicas e/ou privadas, específicas de uma determinada área como, por exemplo, a base de dados American Library Directory; até especificações de produtos e materiais como, por exemplo, a base de dados Thomas Register Online. As

Figura 3. Modelo de Metadados - Base de Dados Marquis Who's Who

NM	Nome biográfico e nome alternativo
FS	Segmento da Edição do Marquis Who's Who
DE, ON	DE= Formação, ON= Identificação da instituição/setor
YB, BP, BC, BS	YB= Data de nascimento, BP= Região de nascimento, BC= Cidade de nascimento, BS= Estado de nascimento
SX	Sexo
DG, NS, YA	DG= Grau de escolaridade, NS= Nome da universidade, YA= Ano em que se graduou
DE	Titulação
CR	Informações profissionais/carreira
DE, PO, CO	DE= Formação, PO= Cargo que ocupa, CO= Nome da empresa
DE/RI	Interesse pessoal/ <i>hobbies</i>
ID/CP	Identificação de atividades cívicas e políticas
ID/HÁ	Identificação de premiações
ID/ME	Identificação de participação associativa
ID/AF	Identificação de filiação político-religiosa
AT	Tipo de endereço: residencial
AD	Endereço
CY	Cidade
ST	Estado
ZP	Cep
AT	Tipo de endereço: comercial
AD	Endereço
CY	Cidade
ST	Estado
ZP	Cep

bases de dados diretório são originárias de publicações impressas como cadastros, anuários, diretórios, guias, etc.

O registro informacional da base de dados diretório Marquis Who's Who (<http://library.dialog.com/bluesheets/html/bl0234.html>) oferece informação sobre pessoas físicas, atuantes nas áreas: industrial; de negócios e finanças; de comunicação e mídia; educacional; governamental; militar; legislativa; judiciária; de saúde; religiosa, de ciência e tecnologia, de esporte; de arte, etc. Atualmente a base possui 790 mil registros de pessoas.

O modelo de metadados da base de dados Marquis Who's Who (Figura 3) compreende 37 campos informacionais que possibilitam o acesso às informações significativas de pessoas de destaque, dentre as quais: um resumo/histórico da carreira; informações sobre a família, religião, ideologia política; *hobbie(s)*; enfim, informações suficientes para conhecer, de modo geral, a trajetória profissional e pessoal da pessoa pesquisada. Constata-se portanto, a utilização de um modelo de metadados bastante completo ao que se propõe a base de dados.

As bases de dados de texto completo possuem dados originais integrais do documento (*full-text*) disponíveis para acesso em meio eletrônico, ou seja, o usuário da base poderá obter o documento completo durante a consulta/pesquisa. Dessa forma, artigos de revistas, normas técnicas, patentes, notícias de jornais, programas de computador etc., podem ser obtidos em sua versão integral.

Os modelos de dados dessas bases de dados são muito diferentes entre si, justamente porque atendem às especificidades da representação e descrição de diferentes estruturas de dados, assim como, da representação de diferentes conteúdos temáticos. As bases de dados de texto completo originam-se, primeiramente, nos jornais diários como, por exemplo, a base de dados Papers e, mais recentemente, nos periódicos eletrônicos, patentes e relatórios de diversas naturezas técnico-científicas.

A base de dados de texto completo U.S. Patents Fulltext cobre a produção de propriedade intelectual/industrial dos Estados Unidos desde 1971 e, atualmente, possui 2 milhões e 245 mil documentos de patentes disponíveis para acesso. Apesar de cobrir a produção de patentes desde 1971, a base disponibiliza textos integrais de patentes americanas, somente a partir de 1974. Além disso, fornece a descrição do desenho (se houver), o desenvolvimento da invenção e o processo legal do pedido de patente (*legal status*).

Pode-se perceber a complexidade do registro informacional da base de dados U.S. Patents Fulltext (<http://library.dialog.com/bluesheets/html/bl0654.html>), justamente porque o modelo de dados representa e descreve todas as informações relevantes do documento de patente. O usuário obtém completude informacional no resultado da pesquisa. A base também fornece informações adicionais, como o desenvolvimento do processo legal do pedido de patente, visto ter agregado valor à informação original, pois esse tipo de informação não está inserido no documento original. Isso demonstra o quanto é importante a base de dados fornecer informações que extrapolem o próprio documento.

O modelo de metadados da base de dados U.S. Patents Fulltext (Figura 4) é composto por 43 campos informacionais. Toda a estrutura textual de um documento de patente é exaustivamente coberta por esses 43 campos de informação. O modelo visa fornecer maior completude informacional no resultado de pesquisa, além de propor melhoria da qualidade de resposta, dependendo da necessidade informacional do usuário.

As bases de dados numéricas podem oferecer desde informações referenciais até textos completos (*full-text*) sobre pesquisas de opinião, de consumo, estatísticas sobre população etc. A estrutura dessas bases de dados também são direcionadas para as especificidades da representação e descrição de diferentes estruturas de dados, assim como da representação de diferentes conteúdos temáticos dos documentos que compõem as bases de dados desse tipo, uma vez que as bases de dados numéricas também trabalham com tipologias documentais bem definidas.

Como exemplo, pode-se citar a base de dados numérica Population Demographics, que oferece dados estatísticos da população americana, baseado no censo realizado naquele país em 1990, totalizando 50.457 registros informacionais. A base de dados possui informações sobre a população, casa/família, renda, educação e ocupação. A base também trabalha com projeções desses dados e oferece estatísticas demográficas, índices sobre o poder de compra, informações industriais, informações de ocupações/empregos.

O registro informacional da base Population Demographics (<http://library.dialog.com/bluesheets/html/bl0581.html>) contempla as informações citadas anteriormente, atendendo as necessidades informacionais dos usuários que necessitam desse tipo de informação. O modelo de dados é bastante completo

Figura 4. Modelo de Metadados - Base de Dados U.S. Patents Fulltext

DT	Tipo de documento (design, processo, etc.)
TI	Título da Patente
PN=	Número do documento de patente
PD, PY	PD= Data de publicação, PY= Ano de publicação
AU, CY, ST, CN	AU= Autoria/Invenção, CY= Cidade, ST= Estado, CN= País - dados referente ao autor
CO, PA,	CO= Empresa/instituição, PA= Autoria institucional da patente
CY, ST, CN, CK	CY= Cidade, ST= Estado, CN= País, CK= Autoria Original - dados referente a instituição
CO, RA, RG, RK	CO= Empresa/instituição autoria, RA= Nova autoria institucional RG= Novo autor, RK= Tipo de responsabilidade
NA	Número de aplicação do documento de patente
AD, AY	AD= Data de aplicação, AY= Ano de aplicação
CL	Classificação de patentes
IC	Classificação Internacional de Patentes
CE	Código do campo de pesquisa
CT	Família de patentes/Citação de patentes
RF	Referências citadas
EX	Nome do examinador
LR	Representante legal
AR	Unidade de artes (desenhos/figuras)
RC	Número e seqüência (dados) do certificado de reexaminação
RE	Número da requisição de reexame
RQ	Requisitor do reexame
AB, TX	AB= Resumo, TX= Texto completo
DW, TX	DW= Síntese da descrição de desenhos/figuras, TX= Texto completo
SU, TX	SU= Resumo do desenvolvimento da invenção, TX = Texto completo
EM, TX	EM= Descrição de vantagens, TX= Texto completo
CM, TX	CM= Contestações, TX= Texto completo

e atende a todas as especificidades das informações coletadas no censo. Além disso, a base de dados oferece análises e projeções de diferentes naturezas como: a) de mercado; b) estatísticas; c) situação local; d) objetivo do mercado; e) planejamento estratégico; f) modelagem de dados; g) territorial; h) econômica.

As informações fornecidas pela Population Demographics, na maioria das vezes, estão divididas em três segmentos: 1) Dados reais do Censo de 1990; 2)

Estimativa para o ano de 1995; 3) Projeção para o ano 2000.

O modelo de metadados da base de dados Population Demographics apresenta 36 campos informacionais (Figura 5) que compreendem as informações nos três segmentos citados anteriormente. O modelo atende às especificidades da estrutura repetitiva de dados e dos conteúdos temáticos uniformes, conforme segue:

Figura 5. Modelo de Metadados - Base de Dados Population Demographics

LV=	Nível Geográfico
ST=	Estado
CX=	Comarca
CE=	Divisão do Censo
DC=, DN=	DC= Código Área Designação de Mercado, DN= Nome Área Designação de Mercado
AA=, AJ=, AT=, BC=	AA= Total de População - Censo 1990 AJ= Total de População - Estimativa 1995 AT= Total de População - Projeção 2000 BC= Média e Porcentagem População
CA=	Média de Idade - Censo 1990
CB=	Média de Idade da População Feminina
CC=	Média de Idade da População Masculina
DA=	População de Raça Branca
DB=	População de Raça Negra
DE=	População de Outras Raças
DF=	População Hispânica
FA=	Renda Familiar até \$9,999
FB=	Renda Familiar de \$10,000 até \$14,999
FC=	Renda Familiar de \$15,000 até \$24,999
FD=	Renda Familiar de \$25,000 até \$34,999
FE=	Renda Familiar de \$35,000 até \$49,999
FF=	Renda Familiar de \$50,000 até \$74,999
FG=	Renda Familiar de \$75,000 até \$99,999
FH=	Renda Familiar de \$100,000 até \$ 149,999
FI=	Renda Familiar acima de \$ 150,000
GA=	Média de Tempo/Anos para Completar a Educação
IA=	Média de Renda Familiar
IB=	Média da Renda Mensal
I0=	Indústria e Ocupação
ES=	Situação de Empregos
HA=, HI=, HO=, HS=	Dados Familiares
MS=	Situação / Estado Civil da População

O modelo de metadados fornece informações e estimativas interessantes por região, ocupação, população, entre outros. Essas informações podem e devem ser usadas pelos diversos segmentos industriais e comerciais daquele país, visando a adequação dos produtos e serviços ao mercado, bem como aos cenários desenhados pelas projeções fornecidas pela base de dados.

As bases de dados financeiras podem oferecer desde informações referenciais até textos completos (*full-text*) sobre organizações diversas como: investimentos, aplicações financeiras, cotações de bolsa de valores, dados e indicadores econômicos, balanços, relatórios anuais etc. Essas bases de dados possuem uma combinação de conteúdos textuais e numéricos que estão presentes, por exemplo, no balanço de uma empresa. O modelo de dados dessas bases são direcionados para as especificidades informacionais de conteúdo. Os dados financeiros, numéricos e estatísticos são coletados no mercado econômico/financeiro; as bases de dados por sua vez trabalham com tipologias documentais bem definidas como, por exemplo, balanços de empresas, relatórios bancários etc.

Assim, a título de exemplo, a base de dados financeira Extel International Financial Cards, oferece cotações, dados financeiros e textuais de uma grande quantidade de empresas internacionais das seguintes regiões: Reino Unido, Europa, Austrália, Hong Kong, Japão, Ásia, América do Norte, Singapura e Malásia. A base de dados possui atualmente 7.675 registros e mantém um histórico de informações cobrindo o período de 5 anos.

A base Extel oferece análises de: a) operação de resultados por região e por linha de negócio; b) informações das diretorias das empresas; c) dados cadastrais das empresas; d) organograma organizacional, incluindo subsidiárias e relacionadas; e) balanço; f) relatórios de lucros e perdas; g) relatórios gerenciais; h) relatórios financeiros; i) informações bancárias e de auditoria; j) detalhes de investimentos em ações; l) aplicação em fundos.

O registro informacional da base Extel International Financial Cards (<http://library.dialog.com/bluesheets/html/bl0500.html>) é complexo e atende todas as especificidades dos conteúdos informacionais anteriormente citados. As informações textuais que a base de dados fornece são obtidas através de diferentes relatórios organizacionais, tais como de marketing, vendas, institucional etc.

O modelo de metadados da base de dados Extel International Financial Cards é complexo e exaustivo, no que diz respeito à completude das informações

financeiras, fornecidas aos seus usuários. O detalhamento das informações, a sistematização dos dados, enfim, toda a estrutura do modelo de metadados (Figura 6) atende às necessidades de qualquer usuário que faça o acesso em busca de informações financeiras de uma determinada organização.

Os 54 campos informacionais que compõem o modelo de metadados da base de dados, demonstram a exaustividade com que a descrição dos dados e a representação dos conteúdos são tratados na base. A consistência e a confiabilidade da base Extel está relacionada à completude das informações analisadas e disponibilizadas ao usuário.

O número de campos informacionais também aponta para a necessidade de a base de dados cobrir, de forma significativa, todas as possibilidades das informações financeiras, bem como projeções e análises agregadas às informações originais. As bases de dados que agregam valor às informações originais, geralmente tem maior custo de acesso, acenando para uma tendência de mercado da indústria da informação.

As bases de dados eletrônicas (*on-line*) ou ópticas (*off-line / CD-ROM*) objetivam agilizar o acesso às informações; observa-se, entretanto, que um dos fatores de insatisfação dos usuários é a revocação das informações recuperadas, isto é, quando um usuário pesquisa em uma base de dados um determinado assunto de seu interesse, na maioria das vezes, ocorre que menos de 50% do que foi recuperado é realmente utilizado ou é relevante para ele.

Esse fato ocorre porque as bases de dados precisam atentar para diferentes variáveis interdependentes como: objetivo da base de dados; política de indexação, particular para cada instituição; vocabulários controlados; modelos eficientes de metadados; objetivo de uso do usuário etc.

Existem várias formas de acessar e recuperar a informação em bases de dados. A mais simples delas é percorrer, registro por registro, à procura da informação, utilizando um termo designado pelo usuário.

O acesso à informação através das listas enlaçadas unidirecionais é realizado através de um campo-chave para localizar um registro informacional como, por exemplo, o campo de autor. Os registros podem ser organizados de acordo com o campo-chave ou organizados independentemente, mas em ambos os casos emprega-se um sistema de indicadores para representar a organização dos registros na base de dados. Esse tipo de gerenciamento de dados não permite que o acesso e a recuperação da informação se desenvolva em outros campos do registro informacional (Rowley, 1994, p.82-83).

Figura 6. Modelo de Metadados - Base de Dados Extel International Financial Cards

PY=	Último ano dos dados coletados
/CO,GN=	CO= Nome da organização, GN= Endereço oficial registrado
TE=	Número do telefone
CN=, IO=	CN= País, IO= Código ISO
NO=	Número Extel da empresa
SN=	Lista Numérica Oficial Diária da Bolsa de Valores
TC=	Código da empresa TOPIC
VN=	Número / em valores
NU=	Número de registro da companhia
EM=	Número de empregados
ID=	Data de registro
SC=, PC=, /DE	SC= Cógigo de Classificação Secundário Industrial Padronizado PC= Código de Classificação Primário Industrial Padronizado, DE= Descrição Código U.K. SIC
/OT	Registrado
L7=	Cifra de negócios (vendas)
L2=	Taxas pré-fixadas (lucro)
L5=	Taxas tributadas aos membros
LV=	Total de capital - cotas/ações
LW=	Créditos/Fundos correntes
L8=	Total de créditos/fundos
L9=	Total de empréstimos
L0=	Cotas de lucro
LZ=	Cotas de dividendos
LY=	Remuneração de diretores
L6=	Valor da rede de bens
LX=	Dados de balanço / preço ações
L1=	Calendário anual de cotações (alta)
L3=	Calendário anual de cotações (baixa)
L4=	Dados de balanço / capitalização no mercado
DI=	Dividendos em porcentagem
PM=	Margem em porcentagem do lucro
WK=	Capital / horas trabalhadas

TV=	Taxa em porcentagem / transações de capitais
DQ=	Equidade / débitos em porcentagem
PR=	Taxa em porcentagem / gastos
PN=	Lucro por empregado
PE=	Salários /preço
DY=	Dividendos em porcentagem (Bruto)
EY=	Lucros em porcentagem (Bruto)
SG=	Performance de negócios em porcentagem / crescimento de vendas em 1 ano
AG=	Performance dos fundos em porcentagem / crescimento em 1 ano
MK=	Capitalização do mercado para os fundos de acionistas (último ano)
PG=	Performance das taxas pré-fixadas em porcentagem / crescimento das taxas pré-fixadas em 1 ano
N4=	Margem de lucro em porcentagem
N6=	Taxa de gastos em porcentagem
N5=	Margem de lucro por empregado
N2=	Preço / salários
N1=	Salários em porcentagem (Bruto)
N3=	Performance de taxas pré-fixadas em porcentagem / crescimento de taxas pré-fixadas em 1 ano
/SB	Atividades do grupo organizacional e da empresa
/AR	Relatórios e balanço comercial

Outra forma de acesso e recuperação da informação é utilizar o método de arquivo invertido, similar a um índice. Esse tipo de gerenciamento de dados trabalha com dois ou três arquivos separados. Como exemplo, pode-se citar um arquivo de texto que contém os conteúdos dos registros propriamente ditos e um arquivo-índice que proporciona o acesso a esses registros. O arquivo-índice é composto de todos os termos-chave designados pela base que podem ser: palavras-chave, autor, data, título de periódico etc., devendo ser a cada entrada de um novo registro informacional na base de dados, atualizado o arquivo-índice (Rowley, 1994, p.84-85).

Esses arquivos são empregados de modo conjunto durante as buscas numa base de dados. O usuário 'entra' com o termo/palavra-chave e o sistema o procura no arquivo índice. Se o termo/palavra-chave não for encontrado no arquivo índice, o sistema informará que não existem ocorrências do termo/palavra-chave pesquisado. No entanto, se for encontrado, o sistema informará quantas ocorrências existem do termo/palavra-chave na base de dados (Rowley, 1994, p.85).

Os termos/palavras-chave que compõem o arquivo-índice apresentam as seguintes características: a) frequência de ocorrência na base de dados; b) arquivo onde está localizado; c) registro no qual está inserido; d) outras informações de localização, por exemplo, parágrafo/campo, posições das palavras no campo etc. (Rowley, 1994, p.85).

Quando esse método de gerenciamento de dados se utiliza de três arquivos diferenciados, o arquivo intermediário permite que os termos/palavras-chave sejam conferidos rapidamente e o sistema informará rapidamente o número de ocorrências na base de dados. Esse terceiro arquivo geralmente é usado em buscas complexas que envolvem várias 'entradas' de termos/palavras-chave (Rowley, 1994, p.85).

Ainda, segundo Rowley, nos sistemas relacionais as informações são mantidas num conjunto de relações ou tabelas. As linhas da tabela equivalem a registros e, as colunas, a campos. Os dados nas várias relações são interligados mediante uma série de chaves (Rowley, 1994, p.86).

A autora afirma que alguns critérios podem ser úteis na constituição de um arcabouço para examinar a qualidade de bases de dados, entre os quais:

- a) "Cobertura: assunto e tipo de material adequados; abrangência, duplicidade de outros serviços;
- b) Atualidade e frequência das atualizações;
- a) Facilidade de uso e preferência subjetiva do usuário;
- b) Saída, como, por exemplo, conteúdo das preferências; qualidade dos resumos; forma (ficha, microforma, papel); em linha ou fora de linha; extensão da saída;
- c) Linguagem de indexação e variedade de pontos de acesso ou campos pesquisáveis, em termos de desempenho de recuperação almejado;
- d) Custo e quem vai pagar;
- e) Documentação, inclusive sistemas de classificação, tesouros, manuais de treinamento etc.;
- f) Meio de armazenamento e possibilidades de acesso;
- g) Formato de registro e estrutura da base de dados" (Rowley, 1994, p.87-89).

Para Takahashi (2000, p. 61) "...a construção de uma sociedade da informação democrática no Brasil está visceralmente dependente do apoio à pesquisa em tecnologia de produção e comunicação de conteúdos..."

O documento elaborado pelas diversas comissões de trabalho do Programa Sociedade da Informação denominado 'Livro Verde' define como uma de suas metas:

"... criar mecanismos para a produção de conteúdos por parte da comunidade. A produção e o uso de conteúdos que reflitam os interesses e explorem a diversidade da riqueza cultural em todos os espaços, de áreas periféricas e rurais, de minorias e de grupos com interesses afins devem ser apoiados, inclusive como uma contribuição adicional para reduzir as disparidades regionais e sociais, abrindo oportunidade para todos os tipos de registro e difusão de manifestações e idéias" (Takahashi, 2000, p.65).

Ainda no documento o grupo de trabalho 'Conteúdos e Identidade Cultural' acredita ser fundamental para a consolidação da sociedade da informação no país "criar um serviço de informação na

Internet que reuna bases de dados sobre documentos produzidos no País ou sobre o País, decorrentes de atividades científicas, tecnológicas, artísticas, culturais, políticas, econômicas etc." (Takahashi, 2000, p.66).

As estruturas das bases de dados devem ser bem desenhadas para que todo o conhecimento gerado no país possa ser acessado e pesquisado pelos diferentes segmentos produtivos da sociedade. A importância de a sociedade brasileira ter a sua produção de informação sistematizada e disponibilizada é vital para o desenvolvimento científico, tecnológico, social e econômico.

REFERÊNCIAS BIBLIOGRÁFICAS

- GROSSMANN, L. O uso de bases de dados no gerenciamento estratégico empresarial. *São Paulo em Perspectiva*, São Paulo, v.8, n.4, p.95-100, out./dez. 1994.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 1991. 347p.
- ROWLEY, J. *Informática para bibliotecas*. Brasília: Briquet de Lemos, 1994. 307p.
- TAKAHASHI, T. (Org.). *Sociedade da informação no Brasil: livro verde*. Brasília: MCT, 2000. 195p.
- TEIXEIRA, C. M. de S.; SCHIEL, U. A Internet e seu impacto nos processos de recuperação da informação. *Ciência da Informação*, v.26, n.1, p.65-71, jan./abr. 1997.

BIBLIOGRAFIA CONSULTADA

- HARTLEY, J.; SYDES, M.; BLURTON, A. Obtaining information accurately and quickly: are structured abstracts more efficient? *Journal of Information Science*, v.22, n.5, p.349-356, 1996.
- LEE, H. Justifying database normalization: a cost/benefit model. *Information Processing & Management*, v.31, n.1, p.59-67, 1995.
- SPINK, A. Term relevance feedback and mediated database searching: implications for information retrieval practice and systems design. *Information Processing & Management*, v.31, n.2, p.161-171, 1995.
- VALENTIM, M. L. P. Bases de dados e a globalização da informação: estudo de formatos eletrônicos e a qualidade de resposta. IN: SEMINÁRIO SOBRE AUTOMAÇÃO DE BIBLIOTECAS E CENTROS DE DOCUMENTAÇÃO, 6,

1997, Águas de Lindóia. *Anais...* São José dos Campos: INPE, 1997.

_____. Política nacional de informação x produção de bases de dados em C&T. IN: SIMPÓSIO INTERNACIONAL DE BIBLIOTECONOMIA "PROF. DR. PAULO TARCÍSIO MAYRINK", 3, 1999, Marília. *Anais...* Marília: UNESP, 1999. p.289-298

VELLUCCI, S. L. 1998. Metadata. *Annual Review of Information Science and Technology (ARIST)*, Medford, v.33, p.187-222, 1998.

_____. Options for organizing electronic resources: the coexistence of metadata. *Asis Bulletin*, Oct., 1997. Disponível em:

<<http://www.asis.org/Bulletin/Oct-97/vellucci.htm>>. Acesso em: 12 jun. 2000.