BASES DE DADOS NO BRASIL: UM POTENCIAL INEXPLORADO

Murilo Bastos da Cunha
Departamento de Biblioteconomia
Universidade de Brasília
70910 Brasília, DF

RESUMO

Apresenta a evolução do uso de bases de dados no exterior e no Brasil. Comenta e analisa os principais problemas que demandam soluções, tais como: custos, acesso a base de dados, treinamento dos usuários, marketing dos produtos e serviços, diferenças nas linguagens de busca, necessidade de maior cobertura de assuntos pelas bases de dados e o papel da iniciativa privada na emergente indústria de informação.

1 — EVOLUÇÃO HISTÓRICA

Um dos eventos mais importantes ocorridos nas últimas duas décadas, na ampla área chamada indústria da informação, foi o surgimento e o crescimento das bases de dados. Pode-se afirmar, sem sombra de dúvida, que foram tais tecnologias que possibilitaram a atual pujança dessa indústria.

Analisando a literatura sobre o assunto nota-se que não está claro qual foi a primeira base a ser citada. Entretanto, sabe-se que as pioneiras foram criadas nos Estados Unidos e que uma das primeiras foi a base de dados numéricos criada em 1951 pelo Bureau do Censo[1], Entretanto, muitos autores concordam que o fator que realmente provocou a explosão das bases de dados foi a sua utilização por parte dos serviços de indexação e análise bibliográfica, através das aplicações relacionadas com a preparação mais rápida de originais de suas publicações para impressão em fotocomposição. Estando pronta a citação bibliográfica, seu resumo em termos de indexação colocados em suporte por máquina, a passagem para outras aplicações foi um caminho natural.

As bases de dados são fontes de informação computadorizadas que podem ser pesquisadas num modo interativo ou conversacional através de um terminal de computador, telex ou mesmo um microcomputador.

Nas tabelas n° 1, 2 e 3 estão listadas, respectivamente, os principais eventos relativos à evolução histórica das bases de dados, os principais bancos de dados de acesso público do exterior e do Brasil.

Tabela 1 — **Evolução histórica das bases de dados**

BATCH	1951	— base de dados numéricos no Bureau of Census (USA).
	1960	— busca em linha feita no S.D.C/Protosynthex.
	1964	— MEDLARS torna-se operacional, busca de SDI em *batch*.
	1965	— Chemical Abstracts Service disponível em *batch*.
		—12/20 bases de dados disponíveis.
	1967	— Engineering Index começa a vender fitas magnéticas.
	1969	— base de dados MARC da Library of Congress.
		— Biological Abstracts começa a vender fitas magnéticas.
		— Data Resources oferece 3 000 séries de estatística.
	1970	— 50-100 bases de dados disponíveis.
BATCH/ONLINE	1971	— MEDLINE operacional em linha.
		— início do OCLC com 54 bibliotecas cooperantes.
	1972	— início do DIALOG com três bases de dados.
	1973	— início do ORBIT com três bases de dados.
	1974	— *NY Times* disponível em linha.
	1975	— 300 bases de dados disponíveis.
	1977	— início do BRS.
	1980	— 600 bases de dados disponíveis.
		— NEXIS inicia base de dados de texto completo.
		— revolução do microcomputador.
	1981	— tecnologia do videodisco usada na recuperação em linha.
ONLINE	1982	— Knowledge Index (DIALOG) e BRS After Dark são oferecidos ao público em geral com linguagens de busca simples.
		— The Source e Compuserve oferecem "supermercado de bases".
	1983	— criação do STN.
	1984	— 2 400 bases de dados disponíveis.
		— sistemas *front-end* para usuários (Easynet e In-Search).
		— problemas do *downloading*.
	1986	— utilização do CD-ROM.

2 — TIPOS DE BASES DE DADOS

No início dos anos setenta existiam menos de dez bases de dados disponíveis através dos bancos DIALOG e ORBIT. Hoje existem, em todo o mundo, mais de 3 200 disponíveis ao acesso público, hospedadas em mais de quarenta grandes bancos de dados. Essas bases de

Tabela 2 **— Principais bancos de dados no exterior**

BANCO	NATUREZA
DIALOG INFORMATION SERVICE	Mais de 350 bases de dados, a maioria referenciais e bibliográficas.
PERGAMON ORBIT/ INFOLINE	Mais de 80 bases de dados, a maioria referenciais e bibliográficas.
TELESYSTEMS/ QUESTEL	Mais de 40 bases de dados, a maioria delas referenciais e bibliográficas.
STN INTERNATIONAL FIZ Karlsruhe, American Chemical Society e Japan Inf. Center of Sci. & Tech.	Mais de 20 bases de dados, principalmente numéricas, com concentração na área de química.
BIBLIOGRAPHIC RETRIEVAL SERVICES (BRS)	Mais de 80 bases de dados, principalmente referenciais e bibliográficas.
EUROPEAN COMMISSION HOST ORGANIZATION (ECHO)	Mais de 15 bases de dados, com desertores em sete línguas, a maioria delas referenciais.
THE SOURCE	Mais de 70 bases; miscelânea de jogos, correio eletrônico, teleconferência, anúncios, finanças, meteorologia, bibliografias, cinemas, TV, teatro, noticiário etc.
COMPUSERVE	Mais de 50 bases, abrangendo comunicações, finanças, informática, compras, educação, jogos, ciência e medicina.

Tabela 3 — Principais bancos de dados no Brasil

BANCO	NATUREZA
Embrapa/DID	• Produz quatro bases de dados, a sua maioria bibliográficas sobre projetos, tecnologias, teses e publicações produzidas por pesquisadores da Embrapa, além de disseminar 12 bases de dados. • Acesso pessoalmente ou por correspondência.
CIN/CNEN	• Sete bases de dados, contendo informações bibliográficas ou cadastrais sobre patentes, normas técnicas, metais etc., além de disseminar três bases de dados. • Acesso direto ou através da RENPAC.
BIREME	• Duas bases de dados na área médica, ambas bibliográficas • Acesso direto ou através da RENPAC.
IBICT	• Permite o acesso a cinco bases de dados, a sua maioria bibliográficas e referenciais, na área de C&T, além de disseminar outras bases de dados. • Acesso direto ou através da RENPAC.
FGV	• Possui duas bases de dados bibliográficos. • Rede Bibliodata. • Acesso pela RENPAC ou pelo telefone.
Prodasen	• 12 bases de dados, a sua maioria bibliográficas, com concentração nas áreas de direito, legislação e jurisprudência. • Acesso através de LP ou RENPAC.
Serpro/ARUANDA	• 2 bases de dados, tendo como mais importantes as bases sobre cadastros industriais, marcas e patentes. • Acesso direto ou através da RENPAC.
DNPM	• Quatro bases de dados, a maioria bibliográficas e cadastrais, nas áreas de sondagens, geologia etc., além de disseminar cinco bases de dados produzidas pelo CPRM.
CENAGRI	• Quatro bases de dados bibliográficas e referenciais em agricultura. • Acesso pessoal ou por correspondência.

dados são de dois tipos (referenciais e de fonte) e seus conteúdos se referem a praticamente todos os assuntos[2].

2.1 — *BASES DE DADOS REFERENCIAIS*

São aqueles que contêm referências ou informações secundárias que identificam as várias fontes primárias. Estas bases não dão a resposta completa a uma determinada pergunta, mas informam aonde o usuário pode ir para obtê-la. Elas são tipicamente de natureza bibliográfica ou de diretórios.

2.1.1 — *Bases de dados bibliográficos* — contêm referências ou citações bibliográficas, com ou sem resumos, da literatura publicada em periódicos, livros, jornais, relatórios, patentes ou teses (Figura 1). Essas bases de dados são as mais utilizadas pelas bibliotecas, como, por exemplo: PERI no Prodasen e Coffeeline no DIALOG.

2.1.2 — *Bases de dados de diretórios* — são as que contêm referências, com ou sem resumos ou sumários, de pessoas, organizações, projetos de pesquisa, contratos, como, por exemplo, o SELAP (Cadastro de Pesquisa e

DIALOG **INSPEC**

```
3/5/1
  355236   B80016852
  SEISMIC DESIGN QUESTIONS TYPIFY NUCLEAR OBSTACLES
  STRAUSS, S.D.
  POWER (USA)   VOL.123, NO.10   55-6   OCT. 1979   Coden: POWEAD
  Treatment: GENERAL,REVIEW-
  JOURNAL PAPER~
  THE SHOCK WAVE GENERATED BY NRCSS CONCERNS REGARDING EARTHQUAKE
RESISTANCE OF NUCLEAR PLANTS HAS LEFT ITS MARK ON AH ALREADY SHAKY
INDUSTRY
  Descriptors: NUCLEAR POWER STATIONS; EARTHQUAKES; SAFETY
  Identifiers: SHOCK WAVE; EARTHQUAKE RESISTANCE; NUCLEAR PLANTS
  Section Class Codes: B8220, B0160
```

ORBIT

```
-1-
ACCESSION NUMBER        B80016852
TITLE                   SEISMIC DESIGN QUESTIONS TYPIFY NUCLEAR OBSTACLES
AUTHORS                 STRAUSS, S.D.
SOURCE                  POWER (USA) (POWEAD), VOL.123, NO.10, PP.55-6, OCT.
                        1979
DOCUMENT TYPE           J (JOURNAL); GR (GENERAL/REVIEW)
LANGUAGE                ENGLISH
CATEGORY CODES          "3B8220; 3B0160
INDEX TERMS             NUCLEAR POWER STATIONS; EARTHQUAKES; SAFETY
SUPPLEMENTARY TERMS     SHOCK WAVE; EARTHQUAKE RESISTANCE; NUCLEAR PLANTS
ABSTRACT                THE SHOCK WAVE GENERATED BY NRC'S CONCERNS REGARDING
                        EARTHQUAKE RESISTANCE OF NUCLEAR PLANTS HAS LEFT ITS
                        MARK ON AH ALREADY SHAKY INDUSTRY.
```

BRS

```
AN B80016B52.
AU STRAUSS, S. D.
TI SEISMIC DESIGN QUESTIONS TYPIFY NUCLEAR OBSTACLES.
SO POWER (USA). VOL.123, NO.10. 55-6. 0.
CD POWEAD.
LG EH.
YR 79.
DE NUCLEAR-POWER-STATIONS, EARTHQUAKES, SAFETY.
ID SMOCK WAVE, EARTHQUAKE RESISTANCE, NUCLEAR PLANTS.
AB THE SHOCK WAVE GENERATED BY NRC'S CONCERNS REGARDING EARTHQUAKE
   RESISTANCE OF NUCLEAR PLANTS HAS LEFT ITS MARK ON AN ALREADY SHAKY
   INDUSTRY.
CC BB220, B0160.
```

Figura 1 — Base de dados bibliográficos

Desenvolvimento, do CNPq), o CAF (Cadastro da Administração Federal, da Semor) e o Dun & Bradstreet PIB (Figura 2).

Registro de Ejemplo
Directorio de Empresas

0006222

MADEPAN INDUSTRIA COMERCIO IMPORTAÇÃO & EXPORTAÇÃO SA
R MAURICIO CARDOSO 1571
CX POSTAL 51/56/57
CRAVATAI, BRAZIL 94000
PHONE: 881066
051-1007
RS

BUSINESS:		INDL INORGANIC CHEM
PRIMARY SIC:	2819	INDL INORGANIC CHEMS.
SECONDARY SIC:	2869	INDL ORGANIC CHEMS
SECONDARY SIC:	3079	MISCH PLASTICS PDTS.
SECONDARY SIC:	5161	CHEMS & ALLIED PDTS.
YEAR STARDED:		1975
EMPLOYEES TOTAL:		867
SALES (LOCAL CURRENCY):		1.950.000,00
SALES (U.S CURRENCY):		12.800,00

THIS IS:
A SUBSIDIARY
AN IMPORTER & EXPORTER
A PUBLICLY HELD COMPANY

DUNS NUMBER:	89-762-1371
PARENT HAME:	GP INDUSTRIA & COMERCIO SA
PARENT DUNS:	89-758-1377
PARENT CITY:	RIO DE JANEIRO
PARENT STATE/PROVINCE:	RJ
PARENT COUNTRY:	BRAZIL
CHIEF EXECUTIVE:	ANTONIO JOAQUIM DE C-PLAMERES-,PRES

(do DAN PRINCIPAL INTERNATIONAL BUSINESSES, baco 518)

Figura 2 — Base de dados de diretórios

2.2 — *BASES DE DADOS DE FONTES*

São aquelas que contêm a informação completa (ou dados primários), as quais não requerem do usuário ir a outras fontes para obter respostas. Elas são de quatro tipos básicos: numéricas, texto completo, dicionário e gráficas.

2.2.1 — *Bases numéricas* — são as que contêm séries estatísticas ou dados numéricos. Em certos casos, os dados numéricos podem ser estatisticamente manipulados para produzir tabelas ou gráficos. Elas também podem incluir dados textuais. Exemplos dessas bases são o Vídeo Bank da Gazeta Mercantil, a Recuperação de Informações por Tema (SRIT) do IBGE e o FMI do Serpro (Figura 3).

2.2.2 — *Bases de texto completo* — são as que contêm o texto completo de um documento, tais como decisões judiciais, legislação, artigos de periódicos ou de jornais. Elas também podem conter dados numéricos. Exemplos dessas bases: Constituições (CONS) do Prodasen, Harvard Business Review no DIALOG, e as publicações da American Chemical Society no DIALOG e STN.

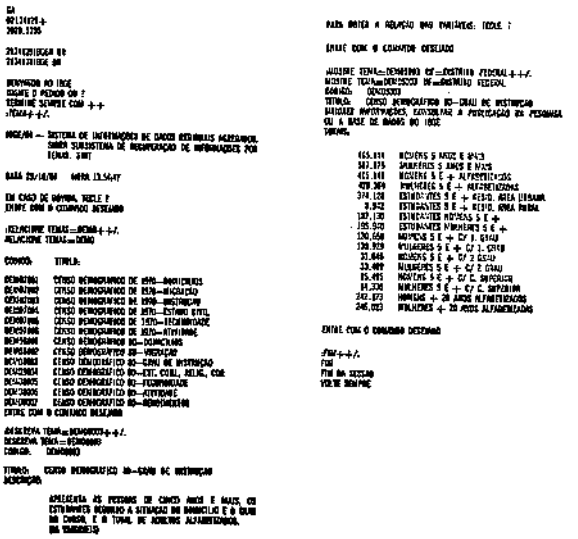


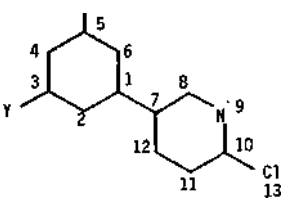
Figura 3 — **Bases numéricas**

2.2.3 — *Bases de dados de dicionários* — são as que contêm manuais ou dicionários, incluindo definições, nomenclatura química, propriedades físicas etc. Exemplos: HEILBRON com fórmulas moleculares de substâncias químicas (Figura 4) e Enciclopédia Kirk-Othmer.

Graphic Display of a Chemical Substructure in the Centre National de l'Information Chimique (CNIC) Online System

USER INPUT:

1-2-3-4-5-6-7-8-9-10-11-12-?
10-13
9 = N 13 = Cl Substitutions = 5,3



CRDS (Chemical Reaction Documentation Service) is available from Derwent

Figura 4 — Base de dados de dicionários

2.2.4 — *Base de dados gráficos* — são as que contêm, sob a forma gráfica, fórmulas de química orgânica, imagens de logotipos, desenhos, figuras e outros. É o mais novo tipo de base de dados e exige *software* de comunicação que tenha capacidade gráfica para fazer recuperações eficientes. Exemplo dessa base é a Trademarkscan (base 226 do DIALOG), que produz graficamente as marcas comerciais patenteadas nos Estados Unidos (Figura 5).

TRADEMARKSCAN® – FEDERAL

03455409
DIMENSIONAL SALON SYSTEMS and Design
US CLASS : 101 [(Advertising and Business)
INTL CLASS : 035 [(Advertising and Business)
STATUS : Registered REG. NO. : 1348455
REG. DATE : July 09, 1985
PUBLISHED : April 30, 1985
GOODS/SERVICES : MANUFACTURER'S REPRESENTATION SERVICES IN THE FIELD OF BEAUTY AND BARBER SALON SUPPLIES-NAMELY, DISTRIBUTING MANUFACTURER'S ADVERTISING MATERIALS IN THE FORM OF BROCHURES, LEAFLETS, SIGNS, PLACARDS, NOVELTY ITEMS AND VIDEO TAPES, AND SELLING PERMANENT WAVE SUPPLIES, HAIR SHAMPOO, HAIR CONDITIONER, HAIR FIXATIVE, HAIR COLORING, HAIR CLIPS, PRODUCT CATALOGS, PAPER BAGS, SWATCH BOOKS, DISPENSERS AND NAIL POLISH TO BEAUTY AND BARBER SALONS
SERIES CODE : 73 SERIAL NO. : 455409
FILED : December 05, 1983
DATE OF FIRST USE : September 01, 1983
ORIGINAL OWNER : ROFFLER OF THE MIDWEST, INC. (MINN. CORPORATION) D.B.A. DIMENSIONAL SALON SERVICES; BURNSVILLE,MN
DISCLAIMER : NO CLAIM IS MADE TO THE EXCLUSIVE RIGHT TO USE "SALON SYSTEMS", APART FROM THE MARK AS SHOWN.
*** Image File *** T3455409.IMG





®EXPATS International Limited,
London, England



®D.O.V.E. Equipment Corp.,
Winfield, MO

Figura 5 — **Base de dados gráficos**

3 —PROBLEMAS QUE DEMANDAM SOLUÇÕES

A busca em linha progrediu de uma fase experimental para tornar-se, pelo menos nos países desenvolvidos, uma ferramenta-padrão nas bibliotecas e centros de informação. Mas, para chegar a este ponto, várias atividades precisaram ser acertadas, melhoradas e implementadas.

A seguir será feita uma análise dos tópicos mais importantes que merecem ser discutidos para uma alteração da situação das bases de dados no cenário brasileiro.

3.1 _ *TREINAMENTO DE USUÁRIO*

O usuário de uma base de dados pode ser visto sob dois aspectos: usuário final e usuário intermediário. Aqui se defrontam duas opiniões divergentes. Uns advogam a necessidade de o usuário possuir um *background* a respeito do assunto objeto da recuperação e que os comandos para a recuperação podem ser facilmente apreendidos. Outro grupo acredita que a recuperação em linha faz parte de um processo maior de ensino de biblioteconomia e que uma graduação e/ou especialização em biblioteconomia é condição importante para se fazerem buscas relevantes.

Independentemente da graduação do usuário, o que realmente importa ter em mente são as qualidades que um bom recuperador em linha deva possuir. Entre elas pode-se mencionar a habilidade de comunicação. Particularmente no que diz respeito à técnica de entrevista, é

necessário conhecer bem o que realmente o interessado precisa ou está demandando. Deve-se incluir também um conhecimento de línguas estrangeiras, principalmente a inglesa, tendo em vista que a maioria das bases estão nessa língua. Trabalho eficiente deve ser também uma característica do recuperador, pois ele será responsável por gastar e/ou controlar recursos da organização ou do cliente. E, por último, mas não menos importante, é a habilidade de datilografar razoavelmente, pois é inadmissível que o técnico fique "catando milho" e gastando preciosos minutos de telecomunicação e outros custos.

Na área de treinamento de usuário de bases de dados, a situação em nosso País deixa a desejar. Das 30 escolas de biblioteconomia que possuem cursos de graduação, até o momento somente a Universidade de Brasília possibilita o contato direto de seus a unos com bases de dados localizadas em computadores fora do campus. Aqui, à semelhança dos países desenvolvidos, os bancos de dados deveriam dar oportunidade para treinamento de futuros bibliotecários (futuros clientes) a preços menores. O ORBIT e o DIALOG, por exemplo, oferecem, há vários anos, acesso a US$ 14/hora para a maioria de suas bases. Sem familiaridade com essa ferramenta de informação, o professor brasileiro não treina seus alunos, e, por conseguinte, os futuros profissionais não conhecem e não utilizam essa tecnologia de informação.

Observem que estou clamando pela ampliação de novos treinandos em nível universitário. Nesse momento, nosso *gap* é ampliado, pois desde 1986 muitas escolas de primeiro grau, nos Estados Unidos, já estão incluindo em suas disciplinas o acesso a banco de dados. Assim, as crianças daquele país já ficam sabendo que as bases de dados são tão importantes como o livro-texto ou a enciclopédia para a confecção de seus trabalhos escolares. A comparação com a nossa realidade é chocante e clama por ações urgentes das autoridades educacionais e da SEI.

Ao se abordar treinamento, deve-se mencionar que a quantidade de cursos oferecidos pelos bancos de dados no Brasil é mínima, chegando às raias do ridículo. Os cursos disponíveis são poucos e a demanda reprimida é grande.

3.2 — *DIFERENÇAS DE COMANDOS*

Numa biblioteca é comum encontrar-se mais de uma enciclopédia. Assim também é necessário fazer contratos com diversos bancos de dados. Aqui se apresenta um grande problema técnico que é a diversidade de comandos utilizados na linguagem de recuperação. Essas diferenças de *softwares* não são uma coisa trivial. Para cada *software* é necessário ler, compreender e armazenar enorme quantidade de documentação. Não existem dúvidas de que uma pessoa será pouco eficiente se tiver que manipular diversos sistemas diferentes.

As figuras n° 6 e 7 mostram as variações de comandos de busca nos principais bancos de dados estrangeiros e brasileiros. O ideal seria que não houvesse diferenças entre eles, ou que, pelo menos, as diferenças fossem mínimas. É claro que aqui entram em jogo pesados investimentos feitos pelos bancos de dados. Mudanças para uma adoção de comandos ou mnemônicos de sistemas de concorrentes não seriam ações simples. Entretanto, considera-se valer a pena mencionar o esforço desenvolvido pelo IBICT para padronizar os comandos de busca utilizados pelos bancos de dados de órbita governamental.

3.3 — *ACESSO ÀS BASES DE DADOS*

Ao consultar os diversos catálogos de bases de dados existentes no exterior, o usuário brasileiro fica com um pouco de inveja da quantidade e diversidade de bases de dados disponíveis lá fora. Será que no Brasil não temos bases de dados? A resposta é: sim e muitas.

Em 1984 a Secretaria Especial de Informática (SEI), após exaustivo levantamento, computou a existência de 1 029 bases de dados em 1982. Em 1986, a mesma SEI, em conjunto com a Secretaria de Modernização da Administração Pública, identificou 390 bases produzidas por 79 entidades públicas[3]. Ao se folhear esse guia, nota-se que muitas bases de dados produzidas por órgãos governamentais são de acesso restrito, dificultando, por conseguinte, o acesso ao público em geral. Tal fato entra em confronto com uma das diretrizes constantes do *Relatório da Comissão Especial sobre Serviços de Consultas a Bases de Dados,* publicado em 1982, em que se afirma ser "direito de cada brasileiro o acesso simples, imediato, amplo e democrático

	DIALOG		QUESTEL		ORBIT		STN	
prompt	?		?		?		=	
conexão c/o sistema	LOGON Nº		LOGON Nº		LOGIN Nº		LOGINID	
conexão c/ base de dados	BEGIN Nº	B	..FILE	FI	FILE	FI	FILE	F
para fazer a pesquisa	SELECT	S	DEFAULT		DEFAULT		SEARCH	S
permite o resultado parcial de uma estratégia de busca	SELECT STEPS	SS						
combina duas estratégias de busca diferentes	COMBINE	C	DEFAULT		DEFAULT		SEARCH	S
lista todas as etapas de pesquisa c/ os seus resultados em uma mesma sessão	DISPLAY	D	..HISTORY	..HI	HISTORY	HI	DISPLAY HISTORY	D HIS
índice da base	EXPAND	E	..DISPLAY	..DI	NEIGHBOUR	NBR	EXPAND	E
mostra os resultados da pesquisa na tela	TYPE	T	..SHOW	..SH	PRINT	PRT	DISPLAY	D
imprime os resultados de pesquisa offline	PRINT	PR	..PRINT	..PR	PRINT OFFLINE	PRT OFF	PRINT	P
finaliza uma sessão de busca	LOGOFF HOLD		..STOP	..ST	STOP	ST	LOG YES	LOGY
desconexão do sistema	LOGOFF		STOP END	ST EN	STOP YES	STOP Y	LOGOFF	
para obter auxílio sobre os comandos	? EXPLAIN		..HELP	..HE	HELP	?	HELP	?
para salvar todas as pesquisas numa sessão de busca	SAVE	SV	..SAVE	..SV	SAVE		SAVE	SAV
para recuperar e executar as pesquisas guardadas	.. EXECUTE	EXS	..EXECUTE	..EX	RECALL		ACTIVATE	ACT

Figura 6 — **Comandos usados no Dialog, Questel, Orbit e STN**

(cont. Figura 6)

OPERADORES LÓGICOS	DIALOG	QUESTEL	ORBIT	STN
para combinar termos	and	and	and	and
para combinar a união de termos	or	or	or	or
para eliminar termos	not	not	not	not
OPERADORES DE PROXIMIDADE	**DIALOG**	**QUESTEL**	**ORBIT**	**STN**
adjacência onde a ordem é relevante	WORD (W)	WORD W	WORD (W)	WORD (W)
adjacência onde a ordem não é relevante	NEAR (N)	D D	--	ADJACENT (A)
permite procurar os termos no mesmo campo, em qualquer ordem	FIELD (F)	FIELD F	FIELD (F)	--
recupera os termos no mesmo subcampo "descritor"	LINK (L)	--	LINK (L)	LINK (L)
recupera termos na mesma sentença	SENTENCE (S)	SENTENCE S	SENTENCE (S)	SENTENCE (S)
TRUNCAMENTOS	**DIALOG**	**QUESTEL**	**ORBIT**	**STN**
truncagem ilimitada	?	+	:	?
truncagem restrita recupera de 0 a 1 caracter	? ?	?	≠	≠
recupera no máximo o nº de caracteres conforme se coloca o sinal	? ?			!
para o meio de termos	?		:	

	IBICT		CIN/CNEN		PRODASEN	SERPRO	
PROMPT	≠		C:				
conexão com a base de dados	DEFINA BASE DE DADOS	DBD	BASE	B	..CHANGE	BASE	DB
procura a expressão de busca	PROCURE	PRO	DEFAULT		..SEARCH	CONSULTE	CS
mostra o resultado de uma busca na tela	MOSTRE CONSULTA	MCO	EXIBE CONSULTA	E	..DISPLAY	IMPRIMA	IM
exibe qualquer documento da base de dados	MOSTRE DOCUMENTO	MDO	EXIBE REFERENCIA	E REF.	.. BROWSE		
lista todas as etapas de busca c/ os seus resultados dentro de uma mesma sessão	LISTE CONSULTAS	LCO	--		--	HISTÓRICO	HI
imprime os resultados de pesquisa	IMPRIMA	IMP	IMPRIMA		..PRINT		
finaliza uma sessão de busca	FINALIZE	FIM	--		--	SUSPENDA	
desconexão do sistema	ENCERRE	ENC	FINALIZE	FIM	..OFF	DESLIGUE	
para obter auxílio sobre os comandos	MOSTRA RESUMO DOS COMANDOS	MRC	?		..HELP		
para salvar todas as pesquisas efetuadas numa sessão	--		--		..SAVE		
OPERADORES LÓGICOS	IBICT		CIN/CNEN		PRODASEN	SERPRO	
para combinar a união de termos	ou	+	ou	+	OR		
para combinar termos	e	&	e	$	pue	e	
para eliminar termos	subtração	—	subtração	:	not		
termos sinônimos					syn		

Figura 7 — **Comandos usados no IBICT, CIN/CNEN, Prodasen e Serpro**

(cont. Figura 7)

OPERADORES DE TRUNCAMENTO	IBICT	CIN/CNEN	PRODASEN	SERPRO
truncagem ilimitada				
truncagem limitada	$	?	/n	?
substitui exatamente um caracter				
OPERADORES DE PROXIMIDADE	IBICT	CIN/CNEN	PRODASEN	SERPRO
adjacência onde a ordem é relevante			adj	
adjacência onde a ordem não é relevante			near	
recupera termos na mesma sentença			with	
recupera termos que pertencem ao mesmo parágrafo numérico			same	

ao conhecimento universal e, em especial, às informações de interesse público, produzidas pelo Estado"[4].

Assim, a tão propalada "abertura democrática" ainda não chegou às bases de dados, existindo muitos obstáculos que deverão ser enfrentados para se chegar à plenitude do direito à informação por parte do cidadão brasileiro.

3.4 _ *DIVERSIFICAÇÃO DE ASSUNTOS*

Como já foi mencionado anteriormente, o número de bases de dados tem crescido bastante nos últimos anos. De cerca de 20, em 1965, temos hoje mais de 3 200 bases de acesso público. Nos países industrializados, as primeiras bases de dados eram de assuntos relativos à Ciência & Tecnologia; aos poucos foram sendo introduzidas as ciências sociais, e agora praticamente todos os assuntos estão incluídos.

No Brasil segue-se esse mesmo padrão de evolução. Segundo levantamento da revista *Info*[5], 50% das bases de dados de acesso público existentes em nosso País estão distribuídas em duas grandes áreas: Ciência & Tecnologia (C&T) e economia e finanças. Os outros assuntos, embora essenciais, não têm ainda sido considerados pelos produtores nacionais. Precisamos, ainda, de bases de dados que incluam assuntos relativos a temas e problemas brasileiros. Por incrível que pareça, não temos nenhuma base de dados de acesso público sobre as duas grandes manias brasileiras: o futebol e o carnaval. Quem sabe não são segmentos de mercado em potencial?

Ao se falar de diversificação de assuntos, é importante citar que uma parcela ainda grande das bases de dados, é produzida no exterior. Na área de ciência e tecnologia o percentual talvez atinja 50%. É claro que, devido ao nível atual de crescimento brasileiro, nessas áreas a dependência de bases estrangeiras perdurará ainda por longo tempo.

É importante também visualizar não só o mercado brasileiro, mas também os países vizinhos, e

quem sabe até os mais distantes. A exemplo dos franceses que, apesar de extremamente ciosos de sua língua materna, já estão vendendo bases de dados com indexação em inglês, os produtores brasileiros, para penetrar no mercado internacional, deverão processar seus dados numa língua de fácil acesso em outra regiões.

3.5 — *CRESCIMENTO DA INICIATIVA PRIVADA*

Estima-se que a indústria de informação nos Estados Unidos envolve negócios de alguns bilhões de dólares. Dentro dessa indústria o segmento de bases de dados ocupa papel importante. O vigor desse segmento não seria possível sem a existência da iniciativa privada, que acreditou em seus produtos e investiu esforços e dinheiro na comercialização da informação legível pelo computador.

Dois grandes bancos de dados, o DIALOG e o ORBIT, tiveram suas criações possibilitadas por recursos governamentais. O DIALOG, por exemplo, teve seu início em 1966, através de contrato assinado para o desenho, programação e implantação de um sistema computadorizado para gerenciar as publicações técnicas da NASA. Já o ORBIT começou suas atividades após ganhar uma concorrência para projetar e implementar um sistema de armazenagem e recuperação de informações para o Departamento de Defesa. Além disso, o ORBIT ainda colaborou na montagem do sistema em linha da National Library of Medicine.

A montagem inicial de bases de dados, e principalmente de bancos de dados, envolve grandes recursos financeiros. No caso brasileiro, o setor privado tem tido pouco auxílio por parte do governo, seja através de recursos a fundo perdido, seja através de linhas especiais de financiamento. As bases de dados mais utilizadas no videotexto da Telesp são produzidas pelo setor privado. Este sobrevive com grandes dificuldades e precisa ser amparado para que se concretize, brevemente, uma real indústria de informação. Semelhante a outros setores da economia brasileira, o setor privado da indústria de informação também precisa de suporte governamental. Exemplos como o da CMA e o da Meca Teleinformática, que já faturam alguns milhões de dólares, precisam ser estimulados. Entretanto, não se advoga, neste trabalho, a saída completa do governo do setor de informação. Tal fato, se ocorresse, poderia representar um perigo para os usuários finais, pois assuntos, mesmo importantes para a comunidade brasileira, que não obtivessem um faturamento mínimo poderiam deixar de ser indexados e difundidos pelas empresas privadas. Assim, a indústria de informação brasileira precisa ter um equilíbrio entre os setores privado e governamental, a fim de que o brasileiro possa ter um acesso democrático e rápido à informação. Parece que *a perestroika* precisa chegar também ao setor de informação, pois em planos governamentais a área privada é timidamente mencionada ou mesmo completamente esquecida.

3.6 — *MARKETING DOS SERVIÇOS*

Considerando o número de bases de dados existentes no Brasil, pode-se afirmar, sem perigo de errar, que o marketing dos produtos e serviços em linha é quase inexistente. A relutância das organizações brasileiras em promover ações de marketing talvez seja causada por uma série de fatores. Entre eles pode-se incluir o fato de a maioria dos bancos de dados estarem vinculados a instituições da órbita pública, setor que, tradicionalmente, não utiliza as técnicas modernas para difusão de seus produtos e serviços. Aqui, entretanto, vale a pena mencionar duas honrosas exceções, o Centro de Informações Nucleares (CIN/CNEN) e, atualmente, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que estão desenvolvendo uma série de ações na promoção de suas atividades em linha.

As ações de marketing são muito mais do que montagem de folhetos coloridos ou aluguel de *stands* em feiras e congressos. A promoção eficaz possui efeito de uma bola de neve, e num certo sentido se alimenta de seu próprio sucesso. O marketing é realmente importante e vital para a saúde do banco de dados. Numa pesquisa[6] realizada em 1981/82 com quatro bancos de dados bibliográficos (BIRENE, BINAGRI, CIN e EMBRAPA), verificou-se que existe uma alta correlação entre o número de usuários de bases de dados e suas atividades de promoção. Entretanto, o marketing não pode ser feito de forma aleatória ou mesmo esporádica. Há necessidade de um plano geral dessa atividade. É importante também ter em mente que na

clientela de informação existem grupos de pessoas com ampla gama de interesses (a noção de segmentação de mercado). A heterogeneidade existe em todas as clientelas de informação, mesmo dentro do perímetro de uma única organização. O esquecimento desse fato pode trazer, muitas vezes, surpresas desagradáveis. Na Embrapa, por exemplo, num levantamento[7] sobre usuários de disseminação seletiva de informação (DSI), verificou-se que 10,5% dos técnicos sabiam da existência de um manual sobre os serviços e que 43% deles nada sabiam sobre essa publicação. Isto apesar de terem sido distribuídas 10 000 cópias da publicação dentro da empresa!

Os bancos de dados não podem ficar numa posição passiva, esperando que as pessoas venham procurá-los por acharem que seus serviços são de qualidade e que esta é evidente para todos. É um sofisma achar que os serviços falam por si mesmos; na verdade, eles sempre precisam ser avaliados, melhorados e divulgados.

Cabe mencionar a importância de o usuário, seja ele o usuário final ou intermerdiário (biblioteca ou *broker* da informação), estar em dia com as atividades correntes e com quaisquer modificações relativas aos serviços. As atualizações dos manuais, mudanças e introduções de novos comandos ou novas caractéristicas sempre devem ser comunicadas aos usuários. É com prazer que se recebem, de forma corrente, os boletins do DIALOG, do ORBIT, do STN, do QUESTEL e do CIN/CNEN, através dos quais se atualizam todos aqueles que fazem buscas para os usuários finais.
É claro que as atividades de promoção envolvem esforços e custos financeiros. Independente das táticas utilizadas, os produtos das bases e de bancos de dados devem observar alguns conselhos[8]:

a) incluir os custos de promoção no orçamento — Esses custos devem constar das despesas normais dos serviços. Um folheto ou um *pôster* para ter sucesso precisa ser bem feito e contar com auxílio profissionais para sua criação e/ou confecção. Isto custa dinheiro. Outras atividades podem envolver demonstrações, envio de disquetes com programas de acesso, *stands* e viagens. Tudo isto deve constar de uma programação orçamentária prévia;

b) conhecer as capacidades e limitações do sistema — Tanto o banco de dados como o intermediário da informação (biblioteca, por exemplo) só devem fazer a promoção dos serviços quando já tiverem conhecimento das capacidades do sistema. O *staff deve* estar suficientemente familiarizado para poder atrair a clientela, caso contrário surpresas desagradáveis poderão ocorrer por ocasião de demonstrações aos usuários;

c) envolver os recuperadores na promoção — O período experimental é o melhor tempo para testar as táticas promocionais. Alguns bancos de dados, como o DIALOG, oferecem, gratuitamente, um certo período de acesso para que os clientes se familiarizem com as novas bases de dados. Muitas vezes esses acessos gratuitos são avaliados pelos bancos e/ou produtores das bases com o intuito de descobrir falhas no novo produto.

37 —CUSTOS

As implicações financeiras da introdução das bases de dados numa organização são grandes. O custo do equipamento necessário para essa atividade está decrescendo rapidamente, mas ainda é inacessível à maioria das micro e médias empresas brasileiras (a mesma analogia se aplica ao setor de bibliotecas).

Existem evidências de que o preço por referência bibliográfica encontrada pode ser muito menor num sistema em linha que aquele alcançado nos métodos manuais. Já em 1978, no Instituto Francês do Petróleo, encontrou-se uma relação de pelo menos 1 para 10 casos mais favoráveis em buscas na área de petroquímica[9]. Todavia, numa pesquisa feita usando base de dados, há um fluxo externo de recursos, enquanto que numa pesquisa manual os custos são todos internos. Assim, pode ocorrer numa universidade ou num centro de pesquisa um técnico despender uma semana numa biblioteca fazendo uma pesquisa manual de literatura no *Chemical Abstracts* sem que alguém questione seu gasto de tempo. Porém, um pequeno custo para alguns minutos de acesso em linha na mesma fonte pode requerer uma considerável justificativa. Em outras palavras, no caso da pesquisa em linha, o dinheiro é enviado para fora da organização a fim de fazer face a despesas com a linha telefônica, a subscrição da base de dados, a

impressão de cada citação bibliográfica e os direitos para o produtos de base de dados[10] O aspecto custo é agravado ainda pela dificuldade de alguns em racionalizar as despesas de acesso às bases de dados, especialmente quando taxas são pagas por serviços que são, num certo sentido, consumidos imediatamente, mesmo quando resultados relevantes não são obtidos.

Os custos do acesso em linha são basicamente formados pelos seguintes componentes:

- Mão-de-obra — salários de recuperador e auxiliares, mais custos trabalhistas adicionais.
- informação — tempo de conexão com o banco de dados, mais o tempo de utilização da base e citações/dados extraídos.
- Reprodução — duplicação, material de consumo, impressão de citações *off-line.*
- Equipamento — terminal, modem, copiadora.
- Espaço — manutenção do espaço físico e dos equipamentos.
- Telecomunicações — taxas da Embratel, impulsos da companhia telefônica estadual e impostos diversos.

Dos componentes acima mencionados, o relativo à mão-de-obra não pesa tanto tendo em vista que os salários pagos no Brasil são relativamente baixos. O equipamento tem tido reduções paulatinas em seu custo. Um microcomputador pessoal custava em 1983 o equivalente a 300 ORTNs (agora OTNs) e, hoje, o mesmo pode ser adquirido por 30% a menos. A reprodução e o espaço físico são componentes que não influenciam muito no cômputo, pois os mesmos também existiriam caso fossem feitas buscas manuais.

O que realmente importa aqui são os custos da informação e das telecomunicações. Os custos da informação, no caso de acesso a bancos de dados no exterior, sofrem acréscimos no mínimo estranhos: 33,33% de imposto de renda sobre pagamento de informações, além de mais de 25% de imposto de operações cambiais sobre pagamento de informações! Isto resulta num agregado de 58,3% a mais sobre a informação obtida no exterior. Esse absurdo agrava, ainda mais, a situação dos usuários, principalmente aqueles envolvidos com pesquisa e desenvolvimento nos pobres centros e universidades brasileiros. A continuar tal excrecência, o *gap* tecnológico entre o Brasil e os países do Primeiro Mundo será cada vez maior.

O componente telecomunicações também é responsável pelo alto custo pago pelo usuário. "E neste contexto, o Brasil aparece como o país de tarifas mais caras do continente. O Interdata, da Embratel, cobra 53 dólares por hora de utilização, o que significa mais do que o dobro cobrado pela Empresa Nacional de Telecomunicações (Entel) do Peru, a segunda em preços. Em terceiro lugar está o Chile, com 14 dólares, em seguida a Argentina, com 13, e por fim o México, com 5 dólares por hora, o mais barato"[11]. Apesar desses últimos números serem de 1986, mesmo com a recente redução ocorrida nas tarifas, o Brasil permanece ainda com taxas elevadas nos preços de telecomunicações. Aqui os usuários precisam se unir para conseguir custos mais acessíveis.

4 — CONCLUSÕES

Considerando o fantástico crescimento no número de bases de dados e também o incremento no número de usuários ocorrido nos países desenvolvidos, e considerando, ainda, que a recuperação em linha no Brasil está em sua infância, espera-se que num futuro não muito remoto esse enorme potencial de informação seja democratizado e esteja disponível a um maior número de brasileiros. No início dos anos noventa possivelmente haverá o *boom* no uso de bases de dados. Além disso, novas bases de dados, cobrindo os mais diversos assuntos, serão produzidas e disseminadas, possibilitando, por conseguinte, que nossos patrícios consigam pegar ainda carona na "terceira onda"!

As evoluções tecnológicas que estão ocorrendo no exterior terão reflexos positivos no Brasil. O uso do microcomputador — existem hoje quase um milhão — deverá ser aumentado e, com certeza, os bancos de dados brasileiros serão

menos burocráticos e mais ágeis na atração de novos clientes e no oferecimento de produtos que realmente sejam importantes para melhorar a qualidade de vida de nossa gente. Não podemos continuar a ser a 8ª economia mundial e ocuparmos, ao mesmo tempo, o 42° lugar em indicadores sociais.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] NEUFELD, M. L & CORNOG, M. Database history: from dinosaurs to compacts discs. *Journal of the American Society for Information Science 37* (4): 183-190. 1986.

[2] WANGER, J & LANDAU, R. N. Nonbibliographic online data base services. *Journal of the American Society for Information Science 31* (3): 170-180, May 1980.

[3] Brasil. Secretaria Especial de Informática. *Diretório de bases de dados.* Brasília, SEI, 1986, 493p.

[4] Brasil. Secretaria Especial da Informática. *Relatório da Comissão Especial sobre serviços de consultas a bases de dados.* Brasília, SEI, 1982. s.p. (princípio 3.2 "b")

[5] Catálogo de bases de dados. *Info,* suplemento de out. 1987, 50p.

[6] CUNHA, Murilo Bastos da. *Base de dados e bibliotecas brasileiras.* Brasília, ABDF, 1984. p. 122.

NOCETTl, Milton A. *Evaluación del programa de disseminación selectiva de información de la EMBRAPA.* Trabalho apresentado ao 6th Interamerican Meeting of Librarians and Documentalists, Santo Domingo, 1981 p. 10.

[8] BAHR, Alice H. Promoting of online services. In: HOOVER, Ryan E. *The information manager's guide to online services.* White Plains, Knowledge Industry, 1980. p. 161-179.

[9] MOUREAU, Magdeleine. Problems and pitfalls in setting up and operating online information services. *Online review 2:* 237-244, 1978.

[10] CUNHA, Murilo B. da. *Base de dados de bibliotecas brasileiras.* Brasília, ABDF, 1984, 224p.

[11] AMÉRICA LATINA ganha rede alternativa. *Info 4* (43); 40-41, agosto 1986.

DATABASES IN BRAZIL: AN UNEXPLORED POTENTIAL

ABSTRACT

The author comments the evolution of the use of databases in several countries, with emphasis on Brazilian utilization. The major problems and solutions related to databases, are discussed, such as: costs, access to databases, user training, marketing of products and services, differences among languages used in the searching process, need to enlarge the range of subjects covered by databases, and, the role of the private sector in the emerging Brazilian information industry.