

# Bases de dados: a metáfora da memória científica

Fernando Sayão

## À GUIA DE INTRODUÇÃO

"Fazemos apelo aos testemunhos para fortalecer ou debilitar, mas também para completar, o que sabemos de um evento do qual já estamos informados de alguma forma, embora muitas circunstâncias nos permaneçam obscuras".

É com essas exatas palavras que Maurice Halbwachs inicia seu livro publicado postumamente, a *Memória Coletiva*. Por uma coincidência extremamente provocante, é também desta forma, ou melhor, é no estado que poderia ser descrito precisamente por essas mesmas palavras de Halbwachs, que um pesquisador, um cientista interroga um banco de dados à procura de informações que insiram seu trabalho de pesquisa na ciência feita pelo seu grupo. Isto é, ele procura um azimute, um quadro de referências que faça com que ele possa reconstruir seu conhecimento sob a luz dos testemunhos dos seus pares e orientar o seu trabalho no sentido estabelecido pela comunidade científica ou acadêmica em que ele está ou deseja estar inserido.

Este estado transitório, no qual se encontra este pesquisador, é chamado por alguns autores da área de ciência da informação de "estado anômalo de conhecimento"<sup>2</sup>. Ele é caracterizado por um alto grau de indefinição em relação ao assunto sobre o qual o pesquisador procura informações. O seu próprio desejo de informação é absolutamente nebuloso, fazendo com que suas interrogações só consigam se realizar durante o ato da busca. O processo de interação com os conhecimentos armazenados na base é que estabelece o foco da questão. A percepção do pesquisador sobre o conhecimento, fatos e dados armazenados nestes meios eletrônicos, articulados com os

seus próprios conhecimentos, recria cenários de conhecimentos mais nítidos.

Wittgenstein<sup>3</sup> diz no seu *Tractatus Logico-Philosophicus* (1922) que "a dúvida, pois, só existe onde existe uma questão, uma questão apenas onde existe uma resposta, e esta somente onde algo pode ser dito". Ninguém interroga uma base de dados sobre o que não conhece. Quando um cientista busca informações em uma base de dados, ele só está querendo validar as suas dúvidas, traduzindo-as por questões autenticadas por seus pares, por questões presentes na memória coletiva de sua tribo. A resposta já existe. Fundamentar sua questão sobre o que já foi estabelecido é uma imposição do método científico, da natureza tribal e cumulativa da ciência, sem o que o pesquisador está condenado ao limbo da rejeição e do esquecimento, e o seu saber ao descrédito. O que ele busca é fundamentalmente enquadrar sua contribuição à ciência comum, consensual. Isto não quer dizer em absoluto que os debates e as controvérsias se aniquilam diante dessa imposição; quer dizer sim que toda discussão, por mais dramática e acalorada que seja, está submetida a um ritual, cuja consulta aos antecedentes, à memória coletiva comum, é uma das etapas privilegiadas.

Quando um pesquisador, diante de um microcomputador ligado a um banco de dados que pode estar em qualquer parte do mundo, vasculha suas estantes eletrônicas à procura de informações que definam, completem ou estabeleçam as fronteiras do seu trabalho de pesquisa, ele repete o mesmo gesto de quem mergulha na memória de seu grupo para reconstruir as lembranças comuns e dessa forma manter íntegra a sua comunidade.

## Resumo

*Traça um paralelo entre as formas de incorporação de conhecimento nas bases de dados internacionais e o conceito de memória coletiva dentro do âmbito da produção científica mundial. Analisa as bases de dados, seus esquemas de representação e recuperação, sua seletividade e as barreiras impostas pelas linguagens de indexação à ciência produzida no Terceiro Mundo. Conclui que as bases de dados são os repositórios dos conhecimentos consensuais gerados pela ciência moderna, constituindo, dessa forma, a memória da ciência oficialmente aceita.*

## Palavras-chave

*Bases de dados; Memória coletiva; Memória científica; Representação de conhecimento; Linguagem de indexação; Indústria da informação.*

Isto nos leva a pensar que as bases de dados, com seus complexos esquemas de representação e de recuperação de informação, que hoje encerram praticamente todos os testemunhos' da ciência moderna, constituem a memória consensual desta ciência; a memória eletrônica de que nenhum cientista pode prescindir para ordenar e reconstruir seus conhecimentos e onde, obrigatoriamente, precisa haver suas contribuições, seus testemunhos inseridos, sob pena de não participar dessa memória coletiva e não ser jamais "lembrado", ou melhor, citado por seus colegas.

Talvez pudéssemos pensar nas bibliotecas especializadas cumprindo este mesmo papel de memória dos saberes científicos. Mas, por mais universal que seja seu acervo, ele não consegue reunir todas as obras de uma área do conhecimento e sua seletividade está baseada em critérios "domésticos", específicos de uma instituição ou um programa. Porém, a diferença mais marcante é que as bases de dados são suportadas por uma tecnologia que permite a simulação, a ilusão de um diálogo, de uma interação em tempo real em uma linguagem que está cada vez mais próxima das linguagens naturais. Existem as interfaces inteligentes, o hipertexto, ajudas sensíveis ao contexto, janelas que criam uma atmosfera conversacional. Esta interação pode ser feita de qualquer lugar onde se tenha um microcomputador e um telefone. É um gesto solitário do cientista à procura dos seus antecedentes; ao passo que a biblioteca interpõe sempre, entre o pesquisador e o acervo, intermediários, vidraças, catálogos, códigos indecifráveis como são as referências bibliográficas.

É sobre este simulacro da memória coletiva científica engendrado pela sociedade pós-industrial - que são as bases de dados e sua relação com os aspectos cumulativo, social e institucional da ciência - que eu gostaria de discutir rapidamente neste texto. Em especial, no que diz respeito à contribuição da memória virtual ao controle e enquadramento da produção científica, que vão desde os limites impostos pelos mecanismos de representação do conhecimento até os desvios ideológicos que conduzem deliberadamente à marginalização de autores, conhecimentos, fatos não identificados com os padrões da ciência oficial.

## A FORMAÇÃO DA MEMÓRIA ELETRÔNICA

O caráter cumulativo da ciência, que se apropria de uma forma rigorosamente seletiva das contribuições de seus pesquisadores, resulta em um corpo de conhecimento baseado no consenso. Este corpo de conhecimento é representado pela literatura técnico-científica, fruto mais óbvio e mais facilmente sujeito à mensuração da atividade científica. São os livros, os artigos de revistas, os trabalhos de congresso, as patentes, portadoras das inovações tecnológicas, os mais autênticos registros da faina diária dos cientistas.

O crescimento vertiginoso da capacidade de armazenamento de dados em meios legíveis por computador - meios magnéticos e, mais recentemente, meios óticos - não foi ainda suficiente para tornar viável o armazenamento generalizado em computador dos conhecimentos gerados pela atividade científica. Esta impossibilidade implica que, para se colocar disponível em sistemas *on-line*, ou seja, sistemas que permitam um certo grau de interação, de conversação em tempo real, é mandatório que o conhecimento registrado na literatura sofra um processo de tradução, de representação, transformando-se em metaconhecimento. Esta tradução é realizada por intermédio de esquemas simbólicos que descrevem tanto a forma física, quanto o conteúdo informacional das obras que estão sendo registradas. O que vai ser armazenado nas grandes bases de dados é uma metáfora da informação original, é o conhecimento virtual, que só existe em função do seu referente, da sua vinculação remota com algum conhecimento real.

A criação dessas representações é factível via linguagens documentárias, que são linguagens artificiais geralmente derivadas da linguagem natural. Essas linguagens são chamadas artificiais no sentido em que não resultam de processo evolutivo e necessitam de regras explícitas para seu uso. Normalmente as linguagens documentárias estabelecem uma relação unívoca entre o termo e o conceito, isto é, entre o signifiante e o significado. Cada termo corresponde a um conceito do sistema de conceitos da área específica com

que se está trabalhando. A construção dessas linguagens é um processo complexo e longo<sup>4</sup>.

As linguagens documentárias são, pela sua própria artificialidade, extremamente redutoras de significado e só podem cobrir conceitos de um domínio específico do conhecimento humano, não havendo, portanto, linguagens documentárias gerais. Elas fazem parte intrínseca dos sistemas de informação, afetando e sendo por eles afetada.

Qualquer linguagem, já se sabe, é uma forma de poder, de dominação. As nossas próprias contradições culturais são um exemplo disso. A linguagem documentária não é exceção. O poder avassalador no sentido da ordenação, da organização que ela exerce sobre a produção literária, especialmente a científica, é chamado pelo semiólogo Umberto Eco de a "ditadura dos resumos". Esta característica é de dramática importância na construção da memória eletrônica da ciência, pois o grau de resolução e entendimento dos conhecimentos que ela apropria está limitado pela capacidade de representação do código. Mas muita atenção: a propriedade redutora de significado dessas linguagens, antes de ser uma deficiência, é o sustentáculo da identidade, do poder de ordenação e classificação, do qual a ciência não pode prescindir, e é, sobretudo, o canal de expressão da ideologia que a ciência suporta. É o seu poder uniformizador que elimina as diferenças desagregadoras, que garante a harmonia na formação das diversas memórias possíveis.

Um exemplo corriqueiro pode deixar mais clara essa relação entre poder uniformizador e controlador da linguagem, a ideologia que ela representa e a formação da memória eletrônica. Apenas 5% da produção científica dos países periféricos está presente nas grandes bases de dados internacionais<sup>5</sup>. Nestas bases, está representada essencialmente a ciência do Primeiro Mundo e os códigos de representação existentes estão voltados para a problemática desses países. Este fato tem grande impacto no armazenamento e recuperação de informações e conhecimentos que estão fora do domínio da ciência e tecnologia primeiromundistas. Por exemplo:

o desejo de incorporar, em uma base de dados internacional sobre fontes de energia, um artigo importante sobre o uso de óleo de dendê e de jojoba produzidos no interior da Bahia como combustível automotivo em substituição ao óleo diesel poderá esbarrar na falta de termos adequados para a representação correta desses óleos e de sua ambientação, resultando em distorções na representação e conseqüentes desvios na recuperação. Isto significa que os próprios limites da linguagem documentária farão com que sejam preservados a uniformidade e o caráter primeiromundista dessa memória eletrônica. Se este artigo chegar um dia a fazer parte dessa base, ele dificilmente será recuperado e, como desdobramento, não será citado por seus pares, ou pelo paralelo que traça este texto, não será "lembrado", pois, insisto, os limites da linguagem determinam o seu esquecimento à medida que o esquema simbólico utilizado é incapaz de expressar com nitidez o conhecimento que ele porta.

Neste ponto, talvez, caiba um parêntese. Aprofundando um grau a mais o paralelo que aqui se traça, quando penso no inverso de tudo isto, ou seja, no acesso a bases de dados internacionais e na recuperação de suas informações, penso sempre que podemos estar, quem sabe, apropriando-nos da memória científica de outras tribos com todas as suas idiossincrasias. Dependendo do uso que se faça das informações contidas nestas bases, esta prática pode constituir um reforço na dependência de paradigmas científicos e tecnológicos, com os quais somos obrigados a conviver por todas as nossas contradições culturais, históricas e políticas. O acesso a estas bases nos deixa sempre a um passo de importação de problemas científicos e tecnológicos estranhos e, muitas vezes, irrelevantes para a nossa realidade terceiromundista, que pode estender nosso esforço de pesquisa para limites irrealis e indesejáveis. Mesmo assim, o acesso a estas informações é de dramática importância para os pesquisadores dos países menos desenvolvidos. Na maioria das vezes, elas são as únicas fontes estruturadas de informações disponíveis, seja pela via *on-line* ou em bases de dados em CD-ROM, posto que ainda

não está plenamente presente, na consciência dos que conduzem a política de informação desses países, a preocupação com a formação de bases de dados nacionais que sigam padrões internacionais, que façam uso de linguagens documentárias adequadas e que sejam a expressão mais completa da ciência e tecnologia praticada por esses países. Essas bases de dados nacionais deveriam reunir os testemunhos da atividade de pesquisa de países ou de regiões, tal como faz a base de dados Lilacs, de forma a reconstruir, para quem as consulta, conhecimentos, cenários, ambientes, fatos e dados pertinentes a um universo próprio. Uma outra questão possível é que a criação de bases de dados bem estruturadas e com um nível de padronização satisfatória talvez seja a forma mais conveniente de tornar visíveis para a comunidade científica internacional a atividade de pesquisa de países ou regiões em desenvolvimento e inseri-la com identidade própria e autenticidade na "grande memória eletrônica".

Entretanto, os processos de exclusão dos conhecimentos gerados nos países periféricos da *mainstream science* são ainda mais variados e estão muito além dos limites impostos pelos códigos de representação. A argumentação sobre a qualidade dos trabalhos, sobre os critérios, ou a falta deles, adotados pelos *referees* das revistas, sobre o caráter regional dessas publicações são ciclicamente utilizados como barreiras para incorporação desses saberes nas memórias eletrônicas. Sem pensar no reducionismo e na ideologia das linguagens de indexação, de uma forma geral, tudo que é escrito em outro idioma que não o inglês é desfavorecido pelos mecanismos de coleta das bases de dados comerciais. Contudo, muitos estudos põem em evidência que uma proporção significativa da pesquisa científica produzida dentro das fronteiras dos países em desenvolvimento, em domínios específicos, mas pertinentes a problemas universais da ciência moderna, são de grande importância tanto para o conjunto desses países, quanto para a ciência em geral. Temas como energia solar, doenças tropicais, agricultura, pecuária são exemplos relevantes que despertam curiosidade científica em qualquer parte do mundo.

Voltando às linguagens documentárias, elas são, em síntese, metalinguagens derivadas da linguagem natural, com semântica e sintaxe própria. Dessa estrutura de representação simbólica, depende, como já foi enfatizado, a formação da memória eletrônica e também das suas possíveis partições. Quero dizer com isto que um único trabalho científico pode ser incorporado em várias memórias, ou seja, ele pode pertencer a várias bases de dados. Os registros em bases de dados distintas das várias leituras possíveis de um documento são viabilizados por códigos de representação específicos e diferenciados, que interpretam este documento mediante regras internas de um sistema de informação qualquer. Esses códigos possibilitam também que esses trabalhos possam ser "lembrados" por diferentes grupos que os valorizem segundo uma ótica própria. É como se um mesmo fato estivesse em memórias coletivas de vários grupos e fosse lembrado de forma distinta em cada um deles.

Suponhamos que um pesquisador escreva um trabalho sobre compreensão de linguagem natural por robôs. Supondo também que esse trabalho ultrapasse todos os filtros de seleção e garanta a sua homologação pela comunidade, ele poderá ser incorporado em uma base de Inteligência Artificial, onde um vocabulário específico fará a representação do seu conteúdo informacional, enfatizando a sua ligação com os problemas dessa área; ao passo que, por um outro código, os linguistas registrarão nas suas bases de dados os fatos sobre lingüística computacional presentes no trabalho e, exagerando um pouco, os pesquisadores da área de robótica fariam uma terceira interpretação.

Como vimos, um trabalho individual, que faz parte do *curriculum vitae*, da biografia de uma criatura, é submetido a várias interpretações e absorvido por memórias de vários grupos. Esta partição é viabilizada pela linguagem. Obviamente, os mecanismos de lembrança, isto é, de recuperação e citação, vão estar submetidos às mesmas regras dessa linguagem.

Mas, se este trabalho não está em nenhuma memória, não pertence a nenhum grupo, não tem existência

reconhecida pela comunidade, ele na verdade não existe. É só um segredo indecifrável na gaveta e no coração de seu autor.

Halbwachs<sup>1</sup> diz que a memória individual é um ponto de vista da memória coletiva. Dentro desta mesma perspectiva, talvez não fosse um exagero dizer que um trabalho científico cumpre o mesmo pressuposto. Ainda sobre o exemplo anterior, poderíamos dizer que um lingüista pode recordar esse trabalho apenas pela gramática especial que ele propôs; um pesquisador de inteligência artificial estará atento às formas de representação de conhecimento necessárias à representação das regras da gramática proposta; alguém da área de ergonomia tem sua atenção desviada pelos fatores que proporcionam melhor interação homem-máquina; os impactos psicossociais serão recordados por um pesquisador em psicologia.

Na verdade, porém, incorporar-se a uma base de dados - na nossa memória eletrônica - é a última etapa de todas por que deve passar uma contribuição à ciência. O ritual se inicia no momento em que o pesquisador determina o escopo de seu trabalho, cujos graus de liberdade estão determinados por constrangimentos sociais, políticos, econômicos e, como veremos adiante, por aspectos mercadológicos. Ao elaborar um projeto de pesquisa, o pesquisador deve estar sensível ao fato de que algumas áreas são oficialmente apoiadas pelos órgãos de fomento à pesquisa e que essas áreas já foram previamente definidas e somente os projetos que estiverem enquadrados nestes planos receberão aval institucional, recursos e financiamentos dos órgãos de apoio à pesquisa. É o caso do Plano Básico de Desenvolvimento Científico e Tecnológico (PADCT), que define no país o que os pesquisadores têm de pesquisar. Estes planos cumprem um papel importante na estruturação, ordenação e homologação da ciência oficial que estarão refletidas nas diversas memórias eletrônicas.

Os autores que logram publicar seus trabalhos em revistas ou em anais de eventos considerados importantes para sua comunidade são aqueles com possibilidade de ter os seus trabalhos incorporados nas bases de dados. Dentro dessa perspectiva, o conhecimento selecionado, representado e registrado nas grandes bases de dados internacionais constitui a documentação sobre a atividade científica oficialmente aceita pela comunidade que a gerou. Essas contribuições receberam o endosso, a homologação dos pares, e receberam, portanto, o direito de pertencer à memória oficial da ciência. Dessa maneira, as bases de dados se constituem na forma mais fiel dos testemunhos dos cientistas. É a esta memória eletrônica que os pesquisadores se dirigem em busca dos referenciais teóricos para as suas atividades.

Um outro fator que tem um impacto determinante na construção da memória eletrônica é que ela constitui um grande negócio. Um negócio que movimenta anualmente milhões de dólares. A informação deixou de ser um bem puramente cultural e transformou-se em bem econômico. Dessa transformação, que modificou totalmente a percepção do valor da informação, apropriou-se o capitalismo, engendrando o que chamamos "indústria da informação", que tem, em outro plano, uma trajetória bastante semelhante à da indústria da cultura no que diz respeito à sua incorporação à estrutura capitalista ou pós-capitalista, como querem alguns, e à sua definitiva transformação em bem comercializável.

Qualquer abordagem do fenômeno da informação a partir da perspectiva econômica implica estar diante de um paradoxo típico do nosso tempo: a informação livre, perene e longeva materializando-se em um bem mutável, temporal, detentor de valor de troca e situado dentro do espaço de relações de um mercado. No contexto da nossa discussão, isto faz com que interesses econômicos e mercadológicos entrem na composição da memória eletrônica.

A construção desta memória, como já foi dito várias vezes, depende fortemente do poder de representação das linguagens documentárias. De acordo com as circunstâncias, elas podem enfatizar fatos e descobertas, como há pouco tempo aconteceu com as descobertas de fatos novos sobre a supercondutividade, ou, mais recentemente, sobre a fecundação de mulheres idosas. Mas também, promover o silêncio, o esquecimento, como vimos no exemplo do óleo de dendê. No entanto, as "zonas de sombras e silêncios", na expressão de Michael Pollak<sup>6</sup>, podem anteceder a linguagem e se instalar como desdobramento de outros problemas, principalmente os ideológicos. Quantos cientistas no mundo inteiro, por suas posições antagônicas aos regimes políticos de suas pátrias, vítimas do totalitarismo, de ditaduras e de intransigências, não tiveram seus trabalhos de pesquisa impedidos de serem divulgados e banidos da memória de sua época?

As bases de dados são, pois, a metáfora da memória da ciência que se pratica hoje. Elas reúnem os testemunhos de pesquisadores com uma linguagem própria, que parece ser mais um instrumento na eterna busca da pedra filosofal da ciência, que é a busca da ordem, do enquadramento, da classificação em um mundo cada vez mais desordenado e mais entrópico.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. HALBWACHS, Maurice. *A memória coletiva*. São Paulo : Vértice, 1990. 189 p.
2. BELKIN, Nicholas J. Anomalous state of knowledges as a basis for information retrieval. *Canadian Journal of Information Science*, v.5, p.133-140, 1980.
3. WITTGENSTEIN, L. *Tractatus logico-philosophicus*, 1961.
4. GOMES, Hagar E. *Manual de elaboração de tesouros monolíngües*. Brasília : PNB, 1990. 78 p.
5. GAILLARD, Jacques. La science du tiers monde est-elle visible? *La Recherche*, n.210, p. 636-640, 1989.
6. POLLAK, M. Memória, esquecimento, silêncio. *Estudos Históricos*, v.2, n.3, p. 3-15, 1989

## Databases: the metaphor of scientific memory

### Abstract

*Comparison between the ways of knowledge incorporation into international databases and the concept of collective memory within the world scientific production is drawn. Barriers imposed on the Third World science by databases selectivity, representation and retrieval scheme as well as indexing languages are analyzed. The conclusions are that data bases are consensual knowledge repositories, generated by modern science and therefore constitute the memory of officially accepted science.*

### Keywords

*Databases; Collective memory; Knowledge representation; Indexing languages, Information industry*

---

### Luís Fernando Sayão

Doutor em ciência da informação.  
Chefe da Divisão de Tecnologia da Informação  
do Centro de Informações Nucleares da  
Comissão Nacional de Energia Nuclear  
(CNEN/CIN)  
Professor-adjunto da Universidade do Rio de  
Janeiro (UNI-RIO)  
e-mail: lsayão@enen.gov.br

---