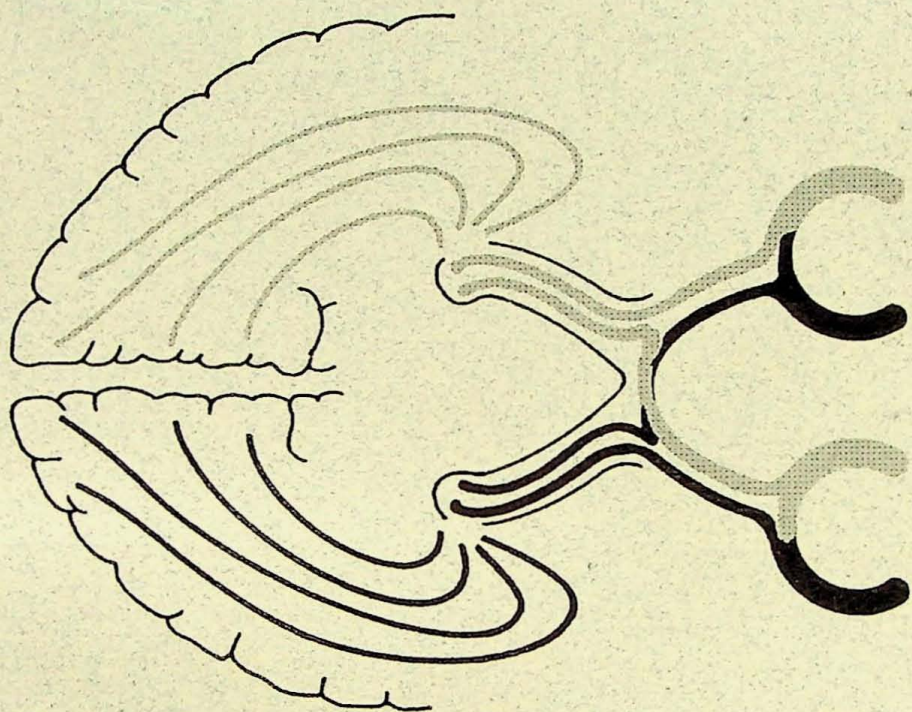


CIÊNCIA COGNITIVA

TEORIA, PESQUISA E APLICAÇÃO

Volume 1, Número 2, p.441-924,
julho/dezembro 1997

ISSN 1415-1472



Editor-chefe: Fernando C. Capovilla

Universidade de São Paulo
Instituto de Psicologia

DESENVOLVENDO HABILIDADES PARA EXECUÇÃO DE BUSCAS EFICIENTES NA BASE DE DADOS *PSYCLIT*

Maria Imaculada Cardoso Sampaio¹
Célia Regina de Oliveira Rosa

*Serviço de Biblioteca e Documentação, Instituto de Psicologia
Universidade de São Paulo*

Resumo: Discute a importância da instrução dirigida a estudantes e pesquisadores para o uso de bases de dados eletrônicas. Apresenta alguns conceitos básicos da estrutura e recursos dos instrumentos de busca, tais como: campos da base de dados, chaves de acesso, descritor, indexação, tesouro, estratégias de busca, truncagem e operadores lógicos. Esclarece as diferenças e semelhanças entre as bases de dados: *PsycLIT*, *PsycINFO* e o índice *Psychological Abstracts*, que constituem instrumentos de busca bibliográfica específicos para a área da psicologia, e oferece exemplos detalhados de estratégias para pesquisa na área. Ressalta que os meios eletrônicos para recuperação da informação são uma realidade indiscutível, mas reforça a necessidade da instrução dirigida ao pesquisador, para que este conheça os conceitos básicos que regem tais fontes e desenvolva as habilidades essenciais para execução de buscas eficientes na base de dados apontada.

Descritores: Bases de dados. Buscas bibliográficas.
Estudantes. Instrução bibliográfica.
Psicologia.

¹ Endereço para correspondência: SBD, PSE, IP-USP, Av. Prof. Mello Moraes, 1721, São Paulo, SP, 05508-900.

O pesquisador conta hoje com uma fantástica variedade de instrumentos que lhe permitem manter-se atualizado quanto à literatura corrente. As facilidades criadas pelas novas tecnologias de acesso à informação só encontram barreiras nas diferentes metodologias utilizadas pelos produtores das bases de dados.

Apesar de toda a tecnologia disponível para a recuperação da informação, poucos são os pesquisadores que conseguem realizar suas buscas bibliográficas na certeza de recuperar documentos relevantes.

Há algum tempo estudiosos da Ciência da Informação discutem a importância da instrução dirigida ao pesquisador no uso de bases de dados eletrônicas, *"users must receive instruction in basic research strategies as well as skills specific to CD-ROM, such as the mechanics of CD-ROM systems."* (Cameron & Hart, 1992, p.239). Mais do que nunca, hoje os pesquisadores hoje devem receber instruções para entender a estrutura das bases de dados, a relação com outros instrumentos de busca, o formato dos registros, a lógica booleana, os princípios usados para indexação da literatura e a importância de planejar a estratégia de busca.

A orientação bibliográfica oferecida pelas bibliotecas precisa ser repensada urgentemente, o pesquisador necessita ser orientado no sentido de adquirir habilidades no uso da tecnologia de recuperação da informação, dominando as fontes eletrônicas e recebendo instruções básicas para que possa ter acesso a esses instrumentos (Murdock, 1995).

O conhecimento de alguns conceitos básicos, como os descritos abaixo, e a compreensão de como a informação é organizada numa base de dados, pode ser eficaz e de grande utilidade para o pesquisador inserir-se no mundo das fontes eletrônicas. Lancaster et al. (1994), demonstraram em seu estudo que um grupo de pesquisadores, após receber um rápido treinamento no uso de uma determinada base de dados e noções gerais de busca bibliográfica no computador, obteve o mesmo resultado final que um bibliotecário experiente.

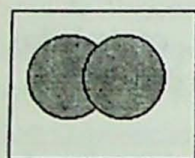
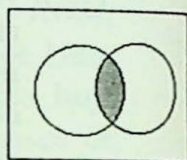
Conceitos básicos:

- **Campos da base de dados:** estruturas definidas para o registro da informação, por exemplo, o campo autor de uma base de dados é o lugar onde são registrados os dados referentes ao nome do autor de uma publicação.
- **Chaves de acesso:** campos definidos para busca numa base de dados.
- **Descritores:** palavras ou grupos de palavras que representam o conteúdo de um documento indexado numa base de dados. O princípio da indexação é que um documento deve ser descrito utilizando-se o termo mais específico possível.
- **Indexação:** processo de análise de um documento para extração dos descritores.
- **Tesouro:** lista de termos (vocabulário controlado) utilizada para se proceder à indexação de um documento.
- **Estratégia de busca:** formulação da combinação de palavras-chave ou descritores que serão utilizados no processo de recuperação da informação. A estratégia de busca é composta, em geral, por palavras ou grupos de palavras combinados entre si através de operadores lógicos.
- **Truncagem:** o símbolo de truncagem (* para algumas bases) permite a recuperação de todos os termos que possuam um radical em comum, por exemplo: digitando-se o termo truncado *inter**, na base de dados *PsycLIT*, serão recuperados os documentos que possuam os indexadores *interview*, *interaction*, *interest*, *intervention*,

interface and *intermediate*. Quando se utiliza a truncagem de um termo corre-se o risco de ampliar demais a busca, recuperando-se literatura irrelevante.

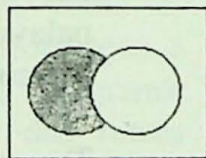
- **Operadores lógicos:** conectores utilizados para ampliar ou restringir a busca. Os operadores lógicos mais utilizados na definição de bases de dados e nos *sites* de busca na *Internet* são os operadores booleanos *or*, *and* e *not*, assim chamados em homenagem ao seu criador John Boyle, matemático britânico que formulou os princípios da lógica booleana.

O operador *and* quando usado, diz ao computador para procurar na base de dados todos os registros que contenham simultaneamente as palavras requeridas; é, matematicamente falando, a intersecção entre dois conjuntos.



O *or* ordena ao computador que traga todos os registros que possuam qualquer uma das palavras pesquisadas; trata-se, portanto, da reunião de dois, ou mais conceitos. O operador *or* une conceitos sinônimos ou equivalentes e pode, desta forma, ampliar a busca.

O *not* indica a exclusão de uma determinada palavra da busca.



O *and* e o *not* são os operadores mais utilizados para se refinar uma pesquisa. Já o *or* amplia o âmbito da busca (Lógica booleana..., 1998).

Em um processo de busca bibliográfica, o conectivo *or* é utilizado quando se deseja rastrear um termo e seus sinônimos ou conceitos equivalentes, por exemplo: *chronic pain or back pain or pain perception or pain thresholds or headache*; serão recuperados todos os registros que possuam um dos termos solicitados. O operador *and* restringe a busca, por ex., *headache and children*; permite que apenas os registros que satisfaçam as duas condições sejam recuperados. O *not* é empregado quando se deseja eliminar um item do tema desejado na busca: *headache not children*, serão recuperados os artigos que tratem do problema da dor de cabeça, porém aqueles que apresentem a palavra crianças serão excluídos da busca.

Conhecendo os instrumentos de busca bibliográfica:

Psychological Abstracts, PsycINFO e PsycLIT

Psychological Abstracts (P.A.): publicação da *American Psychological Association (APA)* é editada desde 1927 com o objetivo de reunir e divulgar a literatura relevante publicada internacionalmente na área da Psicologia. A partir de 1988 deixou de incluir a literatura publicada em outras línguas, que não o inglês. O *P.A.* é atualizado mensalmente, e a partir de 1992 passou a indexar livros e capítulos de livros. Indexa os documentos pelos descritores considerados principais (até o máximo de cinco), não considerando os estágios do desenvolvimento (criança, adolescente, adulto) ou o animal estudado (gato, rato) como cabeçalho de assunto.

PsycINFO: base de dados *online* desenvolvida em 1967 manteve os mesmos objetivos do *Psychological Abstracts*. *PsycINFO* também é atualizada mensalmente, não indexando livros e capítulos de livros e teses na área. Indexa os documentos com até quinze descritores diferentes, permitindo a recuperação da informação por vários tipos de acesso.

PsycLIT: versão em CD-ROM dos dois produtos anteriores, reafirma os objetivos definidos para o *Psychological Abstracts* em 1927. Editada pela primeira vez em 1974, a base de dados *PsycLIT* tem atualização trimestral e indexa livros e capítulos de livros retrospectivamente ao ano de 1987. *PsycLIT* não indexa os relatórios técnicos previstos nas outras duas versões. Como inovação para o ano de 1998, a APA estendeu a cobertura da base ao período de 1967, incluindo dissertações e relatórios técnicos na sua abrangência. O arquivo histórico com referências a partir de 1887 também foi incorporado, reunindo 111 anos da literatura publicada internacionalmente na área da Psicologia. Assim como a versão *online* indexa os documentos com até 15 descritores diferentes, facilitando a recuperação da literatura para o estudioso da área.

PsycINFO e *PsycLIT* continuam a oferecer acesso ao material publicado em outras línguas além do inglês, enquanto o *P.A.* deixou de incluir essas publicações desde 1988.

Na publicação *Psychological Abstracts* um documento pode ser localizado apenas pelo autor ou pelos descritores considerados principais, mas em *PsycINFO* e *PsycLIT* cada palavra do registro constitui chave de acesso para aquele item. Por ex.: palavras do título, nome do periódico, instituição à qual o autor é filiado e muitas outras, inclusive palavras do resumo. Esses recursos permitem às duas versões eletrônicas multiplicar e expandir seu potencial de recuperação, porém exigem maior precisão quando do delineamento da estratégia de busca para que esta seja bem sucedida.

Nas bases de dados eletrônicas *PsycINFO* e *PsycLIT* os artigos são divididos em sessões denominadas "campos", sendo cada campo identificado pelas duas primeiras letras de seu nome:

TI=título do trabalho

AU=autor

IN=instituição a qual o autor é filiado

JN=título do periódico, volume, número, mês e paginação do artigo

IS=ISSN (*International Standard Serial Number*) do periódico

LA=língua do original

PY=ano da publicação

AB=resumo do artigo

KP=frase chave

DE=descriptor

CC=código de classificação

PO=população estudada

AG=faixa etária da população estudada

UD=código atualizado

AN=volume e número da referência no *Psychological Abstracts*

JC=Código do periódico na base

A base de livros conta com outros campos relativos à natureza do material.

Uma busca pode ser limitada a um ou mais campos de acesso, garantindo um alto grau de precisão na recuperação do material. O conhecimento da abreviação dos campos é essencial

para a pesquisa nas fontes eletrônicas. Digitando-se *violence in de* (*de* é a abreviação do campo descritor), serão recuperados os documentos que trazem a palavra *violence* como descritor (cabecinhos de assunto). Acrescentando-se *and English in la* (*la* é a abreviação para a língua na qual o original foi publicado), serão excluídos todos os documentos publicados em outras línguas que não o inglês. Utilizando-se o operador *not social-work in jn* (*jn* é a abreviação para o campo fonte) todos os artigos publicados no periódico *Social Work* serão excluídos da pesquisa. Uma busca com tais características é útil para um pesquisador que já tenha consultado os sumários do periódico *Social Work*.

Entendendo a diferença entre palavra-chave e vocabulário controlado

Entender a diferença entre palavra-chave e vocabulário controlado é essencial na formulação e execução de estratégias de busca. A variação na recuperação é considerável, como mostra o exemplo abaixo:

No.	Records	Request
#1	32635	<i>drugs</i>
#2	20737	<i>drugs in de</i>
#3	3267	<i>drugs- in de</i>

A linha 1 lista o número de vezes que a palavra *drugs* foi localizada como palavra-chave, isto é, o número de registros nos quais o termo aparece em alguma parte do registro: campo do título, campo do resumo, campo do descritor ou no campo da frase chave. Esta técnica de recuperação resulta no que em biblioteconomia se chama alta revocação com baixa relevância.

A linha 2 apresenta todos os registros nos quais a palavra *drugs* aparece no campo descritor (*de* é a etiqueta do descritor ou cabeçalho de assunto). O limitador *in de* significa que a palavra ou frase é recuperada quando aparece no campo descritor. *Drugs in de* (sem o hífen) recupera todos os registros com os descritores *drugs* incluindo seus descritores mais específicos: *emetic drugs*, *ganglion blocking drugs*, *narcotic drugs*. Esta técnica limita bem mais a recuperação, mas não é a busca mais refinada que se pode obter.

A linha 3, com o hífen próximo à palavra e seguido do limitador *in de*, indica o número de registros nos quais *drugs* foi assinalado pelo indexador da APA. O hífen depois de um descritor ou entre palavras é a forma de se dizer ao computador que recupere apenas registros nos quais o termo ou frase aparece desta forma. O resultado é uma busca muito mais precisa (Joswick, 1994, p.50).

Palavras e frases do campo descritor são atribuídas de acordo com o vocabulário controlado da APA. O *Thesaurus of Psychological Index Terms* é a lista de termos do vocabulário controlado utilizado na indexação para descrever o conteúdo dos documentos indexados no *P.A.*, *PsycINFO* e *PsycLIT*; é atualizado a cada dois anos, sendo a oitava edição de 1997. A última edição é composta por cerca de 7726 termos, cada qual com sua estrutura hierárquica e termos relacionados. Novos termos são regularmente acrescentados ao Tesouro, permitindo que a indexação atenda à expansão da terminologia na área da psicologia. Pesquisando sem a ajuda do Tesouro corre-se o risco de recuperar muitos registros irrelevantes que podem apenas mencionar o termo, sem tratar especificamente do tema desejado. O uso do descritor também ajuda a eliminar o emprego não adequado da palavra. Pesquisando-se, por exemplo, pelo termo **adoção** são recuperados registros que tratam não apenas da adoção de crianças, mas sobre a adoção de computadores, a adoção de uniformidade nos negócios e, também, adoção de técnicas utilizadas na criação de galinhas. Pesquisando-se *adoption-child in de* eliminam-se todos os registros que não tratam da adoção de crianças (Joswick, 1994, p.50).

Algumas vezes é mais vantajoso pesquisar recorrendo-se às palavras-chave ou termos livres, pois alguns termos aparecem no título ou no resumo e podem ser exatamente o que o pesquisador deseja. *Soap opera* e *midlife crisis* são bons exemplos de termos utilizados na linguagem comum, que ainda não foram incorporados ao Tesouro e são, no entanto, termos úteis para a busca. Palavras-chave são também uma forma de se chegar ao descritor relevante. Um pesquisador pode digitar um termo em linguagem natural e, examinando os primeiros registros, rapidamente localizar o descritor ideal para sua pesquisa. Por exemplo, digitando-se *gender* como palavra-chave, serão recuperados milhares de registros. Verificando-se os primeiros registros chega-se ao descritor *human-sex-differences*, que é um termo muito mais apropriado para o conceito em questão.

Outras vezes não se encontra um descritor preciso para a busca, mas combinando-se palavras-chave com um termo do vocabulário controlado pode-se chegar aos registros do tema desejado. Pesquisando-se por *serial killer* ou *serial murder* recuperam-se apenas 30 registros. Utilizando-se o descritor relacionado (*homicide- in de*) localizam-se 1321 artigos relacionados com assassinos de todos os tipos, não apenas aqueles que cometem homicídios em série. Combinando-se o termo do vocabulário controlado *homicídio com a palavra-chave serial* (*homicide- in de and serial*) consegue-se reduzir o número dos registros para 43 que são documentos sobre assassinatos que, de alguma forma, tratam do problema de homicídios em série. Nesse caso, a combinação dos termos permitiu a recuperação de artigos que discutem o tema *serial killer* embora, não tenham sido indexados na base por esse descritor, inexistente no vocabulário controlado.

Segundo o princípio da indexação, deve-se utilizar o termo mais específico para a representação do conteúdo do documento. Termos relacionados na hierarquia não são utilizados na indexação. Esta é uma idéia simples, porém importante na formulação de estratégias de busca. Um indexador pode atribuir o descritor *mass-media* para um artigo que fale sobre televisão, rádio ou filmes, mas não descreverá com este descritor um artigo

que fale especificamente sobre televisão. Numa busca onde se digitou *mass-media in de* não serão recuperados artigos que falem exclusivamente de televisão, mas sim todos os artigos cujo conceito esteja internalizado. Da mesma forma *narcotic-drugs* não resgatará documentos que tratem somente de heroína, embora a heroína seja um narcótico.

Se a indexação de um documento foi executada no mais alto grau de especificidade possível, a recuperação deve, obrigatoriamente, valer-se do Tesauro para, através do descritor mais específico, resgatar tal documento.

Quando e porque recorrer ao Tesauro Eletrônico

Recorrer ao Tesauro Eletrônico da base *PsycLIT* (tecla F9) é uma alternativa eficiente quando se deseja incluir subitens do assunto desejado. O Tesauro apresenta uma lista permutada que mostra o termo detalhado, ou seja, o descritor com toda a cadeia de termos relacionados, assim como, o termo mais específico e o mais geral. Por exemplo, na busca por *size discrimination* a lista permutada apresenta todos os termos que incluam a palavra *size* (*size constancy, body size, brain size,...*). Selecionando-se a opção **term detail** (que é um recurso da base *PsycLIT*), a relação hierárquica do termo será exibida.

O descritor *marital relations* traz, como termos específicos, os descritores: *marital conflict* e *marital satisfaction*; uma busca, cujo objetivo seja o de trabalhar o aspecto da relação no casamento, não pode deixar de incluir os termos mais específicos da hierarquia, daí a necessidade de se recorrer ao Tesauro para garantir que todos os registros relevantes estejam no processo de busca.

Term detail identifica, também, o ano que o descritor começou a ser utilizado como indexador, o termo que o precede e quais termos não são utilizados para descrição desse conceito.

A opção **explode term** (outro recurso do sistema) aparece no menu quando um **term detail** é exibido na tela. Ao selecionar-se a opção **explode term** o termo e todos os seus termos específicos podem ser agrupados e transferidos para a tela **find** (tela de busca da base *PsycLIT*), interligados pelo operador *or*. Essa alternativa economiza tempo do pesquisador pois reagrupa os termos e otimiza a recuperação, através da busca de um termo com os seus descritores específicos.

Um termo, apesar de ser apenas relacionado, pode ser de interesse para pesquisador e deve, portanto, ser incluído na estratégia de busca; porém a opção **explode term** não transportará o item para a tela de busca. Nesse caso o descritor deverá ser digitado individualmente, utilizando-se o operador *or*.

Delineando uma estratégia de busca

Tema proposto para um levantamento bibliográfico:
"Aspectos Psicológicos da Preferência por Comida Doce entre Crianças e Adolescentes".

Fonte: CD-ROM *PsycLIT – Journal Articles – 1974-1997*.

Identificando-se o assunto a ser pesquisado, o passo seguinte é o desdobramento da questão em duas ou três chaves de conceitos. Essa tarefa exige habilidade para dividir o assunto em seus componentes principais e sua expressão da forma mais resumida possível. (*Behavioral and ...*, 1997).

Pode-se optar pela elaboração de uma estratégia de busca complexa como a descrita abaixo:

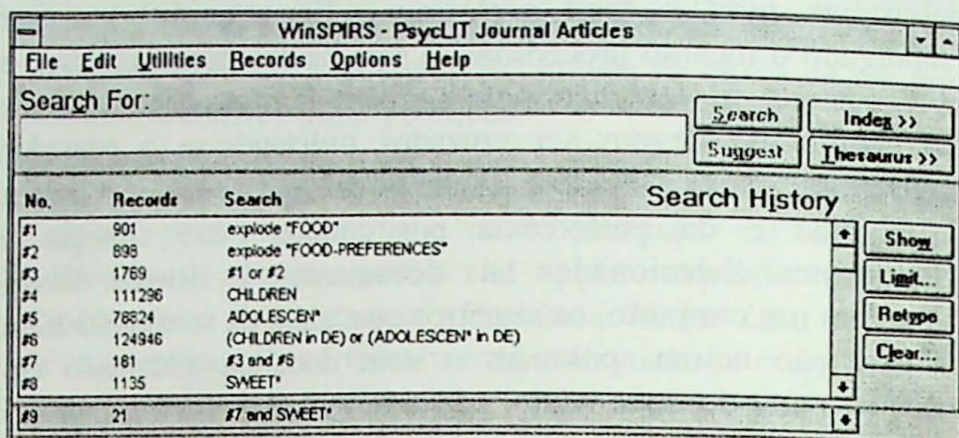
((*food- in de or food-preferences in de*) and (*children in de or adolescen* in de*)) and *sweet**

A ordem dada ao sistema é para que sejam reunidos os dois primeiros descritores em um único conjunto: a somatória dos descritores *food* e *food preferences*. Em seguida, deve ser empregado o mesmo procedimento para os descritores *children* e *adolescents* e *adolescence* [note a ordem de truncagem (*)] e, os dois conjuntos devem ser cruzados utilizando-se o operador booleano *and*. O resultado é a seleção de documentos que tratem da comida e da preferência por comida entre crianças e adolescentes. Selecionados tais documentos, o sistema deverá isolar, em um conjunto, os registros que, além de se enquadrarem na condição acima, possuam o item doce ou adocicado em qualquer parte do documento, pois o termo *sweet* foi truncado (*) e não se limitou a busca a nenhum campo específico. A tela a seguir apresenta o resultado da estratégia executada.

No.	Records	Search
#1	963	FOOD-
#2	898	FOOD-PREFERENCES
#3	111296	CHILDREN
#4	78824	ADOLESCEN*
#5	1135	SWEET*
#6	21	((((FOOD- in DE) or (FOOD-PREFERENCES in DE)) and ((CHILDREN in DE) or (ADOLESCEN* in DE))) and SWEET*

Esta opção de busca requer do pesquisador habilidade no uso do Tesauro; para encontrar os descritores, e conhecimento de delineamento de estratégias de busca, inclusive com utilização de parênteses, para ordenar ao sistema qual condição deve ser estabelecida, e em que ordem.

Para o pesquisador não tão experiente com os recursos oferecidos pelo sistema, a mesma busca pode ser efetuada de acordo com a ilustração a seguir.



Utilizando-se a opção **explode term** para o descritor *food*, selecionou-se também o descritor *food preferences*. O operador *or* foi automaticamente utilizado pelo sistema. Em seguida os descritores *children* e *adolescen** são selecionados formando um outro conjunto. O resultado é submetido ao cruzamento com o radical *sweet* e seus derivados, obtendo-se a recuperação de 21 registros nas condições exigidas.

Tal possibilidade de busca facilita o uso dos recursos do sistema quando não se tem muita familiaridade com a metodologia da base, não requerendo o emprego de estratégias de busca complexas.

Considerações finais

Os meios eletrônicos para recuperação da informação se apresentam como uma realidade indiscutível, as fontes bibliográficas impressas estão sendo substituídas por versões em *CD-ROM* e *online*, e a tendência é que saiam de circulação, dando lugar às fontes eletrônicas, pois são ilimitados os benefícios

extraídos de buscas bibliográficas nesses suportes. Entretanto, o pesquisador só se beneficiará desses recursos obtendo o conhecimento básico dos conceitos que regem tais fontes e desenvolvendo habilidades no uso desses instrumentos. Os princípios nos quais a informação é organizada, tanto em bases específicas, quanto nos sites de buscas da *Web*, são fundamentados na mesma lógica, e o pesquisador que adquirir habilidades gerais para pesquisar em uma determinada base, torna-se auto-suficiente para fazer uso da maioria das fontes eletrônicas disponíveis para recuperação da informação.

SAMPAIO, M.I.C., ROSA, C.R.O. Guidelines for developing effective research skills using *PsycLit* database. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, São Paulo, v.1, n.2, p.629-44, 1997.

Abstract: The paper discusses the importance of training psychology students to use electronic databases on their own. It presents some basic concepts regarding both structure and resources provided by search tools such as: database fields, access keys, descriptors, indexing, thesaurus, search strategies, truncated words and logical operators. It explains some differences and similarities among the main instruments of bibliographic search in the area of psychology, i.e., the databases *PsycLIT* (CD-ROM), *PsycINFO* (online) and *Psychological Abstracts* (bibliographic index). It also offers examples of detailed strategies for searching a theme and emphasizes that information retrieval through electronic media is an unquestionable reality. However, it stresses the importance of providing students with instruction in order to make them familiar with the basic concepts that rule such sources and to the develop the necessary skills performing efficient searches in those databases.

Index terms: Bibliographic instruction. Bibliographic searches. Databases. Psychology. Students.

REFERÊNCIAS BIBLIOGRÁFICAS

- BEHAVIORAL and psychological aspects of food and eating. *PsycINFO News*, v.17, n.4, p.3-4, 1997.
- CAMERON, L.; HART, J. Assessment of *PsycLIT* competence, attitudes, and instructional methods. *Teaching of Psychology*, v.19, n.4, p.239-42, 1992.
- JOSWICK, K.E. Getting the most from *PsycLIT*: recommendations for searching. *Teaching of Psychology*, v.21, n.1, p.49-50, 1994.
- LANCASTER, F.W.; ELZY, C.; ZETER, M.J.; METZLER, L.; LOW, Y.M. Searching databases on CD-ROM – comparasion of the results end-user searching with results from 2 models of searching by skilled intermediaries. *RQ*, v.33, n.3, p.370-86, 1994.
- LÓGICA booleana? Que é isso? *Internet World*, v.2, n.24, p.74, 1998.
- MURDOCK, J. Re-engineering bibliographic instruction: the real task of information literacy. *Bulletin of the American Society for Information Science*, v.21, n.3, p.26-7, 1995.
- THESAURUS of psychological index terms. Ed. Alvin Walter Jr. 8.ed. Washington, DC, Americam Psychological Association, 1997.