

# **Novo modelo de processamento e difusão da produção científica, técnica e artística da Universidade de Brasília<sup>1</sup>**

**Jaime Robredo<sup>2</sup> e Leandro Rodor de Oliveira<sup>3</sup>**

## **1. INTRODUÇÃO**

A principal preocupação do usuário, qualquer que ele seja, quando consulta uma base de dados para obter respostas às suas necessidades de informação, é a relevância – ou pertinência – dos resultados (GREISDORF, 2000). Embora o esquema clássico, reproduzido na *Figura 1*, pareça indicar que todo o problema se reduz a conseguir a harmonização entre os termos utilizados para indexar os conceitos significativos dos documentos e aqueles utilizados para formular as estratégias de interrogação das bases de dados (ROBREDO; CUNHA, 1986), a realidade não se apresenta tão simplesmente assim. Com efeito, o crescente uso da Internet para acessar as mais diversas bases de dados, incluindo os acervos das bibliotecas virtuais e outras múltiplas fontes de informações disponíveis por intermédio dos poderosos motores de busca de *AltaVista*, *Yahoo*, *Google*, *MetaCrawler*, *alltheweb*, etc., eliminam praticamente a intermediação dos profissionais da informação e tornam inviável todo uso de vocabulário(s) controlado(s) ou tesouros (quais?, quantos?, onde?, como?) pelo usuário. Isso sem falar dos custos cada vez mais elevados da indexação chamada manual ou controlada (não

---

<sup>1</sup> O presente trabalho é resultado de uma parceria entre o Decanato de Pesquisa e Pós-graduação (DPP), da Universidade de Brasília (UnB) e a SSRR Informações Consultoria e Projetos Ltda, visando o desenvolvimento e implementação de um novo modelo de processamento e difusão da produção científica, técnica e artística da Universidade.

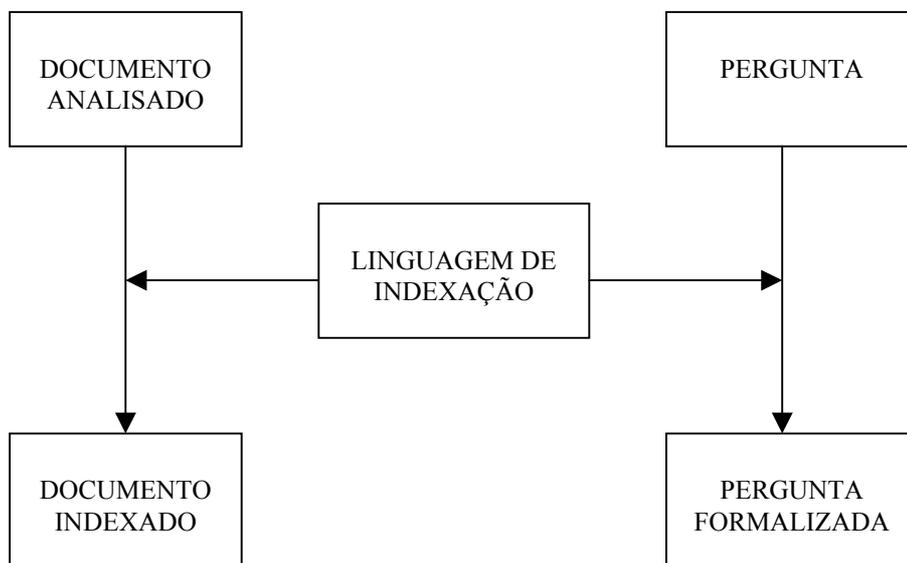
<sup>2</sup> Doutor em Ciências. Pesquisador Associado Senior, Departamento de Ciência da Informação e Documentação (CID), da Universidade de Brasília (UnB); Consultor da SSRR Informações Consultoria e Projetos Ltda, Brasília DF. E-mail: ssrrinfo@uol.com.br.

<sup>3</sup> Bibliotecário, Consultor da SSRR Informações Consultoria e Projetos Ltda, Brasília DF. E-mail: ssrrinfo@uol.com.br.

deveria ser chamada de indexação intelectual?), que a restringem à indexação de bases de dados altamente especializadas, normalmente acessíveis mediante assinatura, cujo jargão somente dominam os técnicos, especialistas e cientistas dos domínios correspondentes da ciência e/ou da tecnologia. Assim, a bibliografia sobre novas abordagens da representação dos conteúdos, com ênfase crescente no auxílio da ciência e técnica da computação, junto com as abordagens cognitivas de estudo do comportamento do usuário – focalizando seus diversos tipos, culturas e níveis de conhecimento –, frente aos sistemas automatizados, avoluma-se de forma impressionante. Por isso, nos limitamos a apresentar somente algumas referências sobre o assunto (ELLIS, 1992; LAGER, 1996; BATES, 1998; CHALMERS, 1999; SOERGEL, 1999; PALMER, 1999; WOLFRAM, D., 1999; WOLFRAM, 2000; WILSON, 2000; SPINK. *et al.*, 2001; BENKING, 2002), que talvez não sejam as mais adequadas ou mais atualizadas (as coisas evoluem tão rapidamente!), de forma que outros estudiosos certamente sugeririam substituí-las ou complementá-las, mas que certamente servem de amostragem e podem permitir aos interessados, a partir das suas copiosas bibliografias, e, sobretudo, com o auxílio de eventuais buscas na Internet, chegar o cerne do objeto de seus próprios interesses. Neste trabalho, estudaremos sucessivamente, de forma sucinta: *a)* alguns dos diversos fatores que podem favorecer ou prejudicar a qualidade (ou fidelidade) da indexação, *b)* algumas características que deve possuir um motor de busca razoavelmente bom; *c)* a conjugação desses dois elementos para tratar de otimizar a harmonização das linguagens de indexação e de busca.

Uma breve apresentação de alguns exemplos ilustrativos permitirá acompanhar os principais passos do processo de desenvolvimento e implementação do modelo, assim

como os resultados alcançados, que abrem interessantes perspectivas para futuros aprimoramentos.



*Figura 1. – A linguagem de indexação é o elo entre o conteúdo dos documentos e a linguagem das perguntas.*

## **2. A INDEXAÇÃO**

Que a qualidade da indexação manual, quando feita por especialistas que se utilizem de tesouros, dicionários, etc., sem pressões de tempo para dirimir dúvidas e, ainda, sem erros na digitação dos descritores, seja superior à que se pode obter com procedimentos de indexação assistida pelo computador ou totalmente automatizada continua sendo tema de controvérsia. Entretanto, a julgar pelos rumos para os quais aponta a literatura das últimas décadas, a indexação manual parece perder força progressivamente. E isso por várias razões:

- o custo dos indexadores especializados, cada vez mais difíceis de serem encontrados, inviabiliza a sua contratação;
- o tempo necessário para realizar uma criteriosa indexação manual é incompatível com a exigência generalizada de tornar acessíveis, quase que instantaneamente, enormes volumes de dados e informações;
- o uso de tesouros e outros vocabulários controlados, geralmente usados na indexação manual, privilegiam o uso de determinados termos em detrimento de seus

sinônimos, o que se traduz em perda de informação quando um usuário interroga a base usando um termo sinônimo que representa o conceito procurado tão bem quanto o descritor autorizado;

- os tesouros e dicionários utilizados na indexação manual raramente estão atualizados e, como conseqüência, muitos documentos indexados entre uma atualização e a seguinte serão irremediavelmente perdidos por não terem sido indexados com os novos termos que constantemente são incorporados aos vocabulários especializados;

Uma interessante obra sobre indexação e elaboração de resumos, por métodos automáticos, foi publicada por O'CONNOR em 1996, e sua leitura continua sendo útil. A variedade de sistemas de indexação automática performantes é atualmente considerável. Deixando de lado os sistemas baseados na lingüística computacional, que despertaram grandes expectativas nas décadas passadas, mas que requerem enormes esforços intelectuais para desenvolver as gramáticas correspondentes, que os tornam caros, vamos dirigir nossa atenção para os sistemas de filtragem das palavras vazias de significado, utilizando listas de termos que carecem de qualquer conteúdo conceitual (preposições, advérbios, conjunções, verbos auxiliares, etc.) e que, portanto devem ser desconsiderados no processo de identificação automática de 'descritores' (UNESCO, 1993; UNESCO, 1997). Algumas de suas vantagens seriam:

- rapidez;
- eliminação da digitação dos 'descritores';
- respeito da grafia original das palavras do texto;
- representação dos conceitos por todas as palavras sinônimas utilizadas no texto original, o que se traduz pela multiplicidade das opções de escolha dos termos no processo de busca;

Algumas desvantagens aparentes seriam:

- as listas de palavras vazias de significado deve ser estabelecida previamente e, como no caso dos tesouros e dicionários, devem ser adaptadas aos domínios em que serão aplicadas, só que estabelecer listas de palavras vazias é muito menos dispendioso e muito mais simples e rápido do que organizar um tesouro;
- os sistemas de indexação automática baseados no princípio de filtragem de palavras vazias rendem 'descritores' simples (palavras) e não expressões compostas o que exige, no momento da busca, o uso de conectivos (v. Seção seguinte);

- um mesmo radical pode dar lugar a palavras flexionadas ou com várias desinências, de forma que um mesmo conceito pode ser representado em forma verbal flexionada, ou substantivada ou adjetival com todas as diferenças possíveis de gênero e número, o que obrigará a fazer uso da truncagem dos termos, no momento de formular as estratégias de busca;

O uso de uma dupla filtragem do texto, consistente, como indicado acima, na exclusão das palavras do texto identificadas na lista de palavras vazias, seguida de uma segunda filtragem de palavras flexionáveis ou com desinências variáveis, sem valor como ‘descritores’, a partir de uma nova lista, esta de radicais não significativos (ROBREDO, 1995).

A intervenção humana, nos sistemas de indexação assistida pelo computador, permite ‘editar’ o texto original de forma a delimitar expressões compostas, mediante sinais ou signos predefinidos. Pode-se citar, por exemplo, <biblioteca pública>, no MicroISIS (UNESCO, 1993; UNESCO, 1997) e variações dele derivadas, e biblioteca\_pública, no InfoDoc (ROBREDO, 1995).

Na procura do tudo automático, os recursos da indexação assistida por computador vão perdendo fôlego (custo, tempo, subjetividade, etc.) frente às novas possibilidades abertas pelas novas correntes da pesquisa e de seus desenvolvimentos. A título de exemplo, citaremos somente duas: as aplicações infométricas (LARSON, 1996) e as baseadas no uso das linguagens hipertexto de tipo dinâmico (KALFATOVIC, 2000). Os métodos infométricos que permitem determinar a probabilidade de co-ocorrência de dois termos considerados como ‘descritores’, no processo de indexação automática dos registros da base de dados, são a base de numerosas aplicações práticas (RUGE, 1992; ROBREDO; CUNHA, 1998), dentre as quais cabe destacar duas: uma que pode contribuir para melhorar o nível de especificidade da indexação por filtragem, identificando, no conjunto de termos que apresentam elevada probabilidade de co-ocorrência, se determinados pares de termos se apresentam no texto adjacentes ou próximos e, então, o algoritmo de indexação pode formar automaticamente determinadas expressões compostas significativas; outra variante dessa aplicação infométrica, pode fazer com que a seleção automática de um ‘descriptor’ seja seguida do registro automático como ‘descritores’ de todos os termos sinônimos, mesmo se não se encontram no texto. Nas aplicações das linguagens hipertexto dinâmicas (HTML

dinâmico, DHTML, XML, etc.), merece destaque a possibilidade de associação automática (*links*) de vários documentos ou de partes (capítulos, seções, etc.) de documentos volumosos, a partir de um léxico, que reúne termos pré-estabelecidos, de forma que documentos que pertencem a um mesmo domínio ou sub-domínio encontram-se interligados, o que pode facilitar enormemente a organização dos dados e informações na base de dados, com a conseguinte melhora da recuperação (CLEVER PROJECT, 1999).

### 3. O MOTOR DE BUSCA

O motor de busca pode ser considerado bom quando permitir a obtenção de bons resultados na recuperação, no sentido de alto índice de relevância – ou seja, baixo índice de ruído – dos registros recuperados, e elevado grau de exaustividade, o que significa a menor quantidade possível de registros pertinentes existentes na base que não foram recuperados. Se o índice de relevância, e, conseqüentemente, o índice de ruído podem ser calculados facilmente, o mesmo não ocorre com o índice de exaustividade, pois para poder calculá-lo seria necessário conhecer quantos registros pertinentes existem na base, e isso ninguém sabe. Para o que nos interessa, nos limitar-nos-emos a lembrar alguns recursos que podem ser chamados de ‘auxílios à recuperação’ (Robredo, 1986), e que, quando utilizados criteriosamente e, sobretudo, de forma coerente com recursos implementados para operacionalizar o sistema de indexação automática, permitem aumentar os índices de pertinência dos resultados obtidos na recuperação. Pode-se citar:

- O uso dos conectivos (ou operadores booleanos) E (and), OU (or) e NÃO (and not):
  - o conectivo E [**A** and **B**] recupera os registros que possuem simultaneamente os descritores **A** e **B**;
  - o conectivo OU [**A** or **B**] recupera os registros que possuem um dos dois descritores **A** ou **B**, ou os dois;

- o conectivo NÃO [A and not B] recupera os registros que possuem o descritor **A**, excluindo os que possuem os descritores **A** e **B** simultaneamente.

A expressão de busca [A or B] recupera o maior número de registros (normalmente, mais que as buscas que procuram por **A** ou **B** sozinhos); a expressão [A and B] aumenta a precisão da pesquisa e recuperará sempre menos registros que as expressões [A or B], ou [A] ou [B] (a única exceção ocorre no caso – muito improvável em grandes base de dados – em que somente existe na base o descritor **A** e não o **B**, ou vice-versa); a expressão [A and not B], recuperará, normalmente, menos registros que a expressão [A and B] e, naturalmente, menos que a expressão [A or B], e também menos que [A] ou [B], com única exceção do caso – muito improvável em grandes base de dados – em que **B** não ocorra nunca simultaneamente com **A**, no mesmo registro.

- A truncagem a direita que permite, conservando o radical, recuperar todos os termos que possuem esse radical, qualquer que seja a flexão ou a desinência. Por exemplo: a expressão de busca **ANALI\$** recuperará os registros que contêm os termos:

**ANALI SA**  
**ANALI SADA**  
**ANALI SADAS**  
**ANALI SADO**  
**ANALI SADOS**  
**ANALI SAM**  
**ANALI SA**  
**ANALI SARAM**  
**ANALI SE**  
**ANÁLI SE**  
**ANALI SEM**  
**ANÁLI SES**  
**ANALÍ TICO**  
**ANALÍ TICOS**  
**ANALÍ TICA**  
**ANALÍ TICAS, etc.**

e também expressões como:

**ANÁLI** SE CLÍNICA  
**ANÁLI** SE TERMICA DIFERENCIAL  
**ANÁLI** SE QUÍMICA QUANTITATIVA, etc.

Amplio demais? Sim, pode ser, mas como as formas verbais de um radical significativo são certamente muito menos freqüentes que as formas nominais ou adjetivais do mesmo radical, o provável aumento do índice de ruído não é, na prática, significativo. No caso de uso combinado da truncagem com operadores booleanos, as vantagens podem ser realmente importantes. Por exemplo, uma busca formulada como **BIBLIOT\$ and PUBL\$**, vai permitir recuperar registros que falam de **BIBLIOTECA PÚBLICA, BIBLIOTECAS PÚBLICAS**, “...novas **bibliotecas** para **públicos** diversificados...”, “...aumenta a freqüência do **público** na **Biblioteca** Municipal em horário de almoço...”, etc., e renderá, em uma única busca, um número de registros significativos que, de outra forma somente poderiam ser localizados com repetidas buscas que levassem em conta as numerosas formas de representar o mesmo conceito.

- A indicação de adjacência ou proximidade que permite associar, na formulação da busca, dois termos que se encontram juntos (adjacência), ou separados por um certo número de palavras (proximidade). Existem várias formas de sintaxe para implementar os algoritmos correspondentes. No caso de adjacência, é comum utilizar o conectivo ADJ. Assim, a expressão [A adj B] significa que desejamos recuperar registros nos quais A e B se encontrem juntas, seja na forma **AB** ou **BA**. No caso de proximidade, pode-se indicar o número de palavras que separam os dois termos que se deseja associar, ou os limites de um intervalo. As expressões **A(1)B** e **A(3)B** significam buscas em que desejamos que os termos A e B se encontrem

separados, respectivamente, por uma ou por três palavras, e a expressão **A(2,4)B** significa que aceitamos qualquer resultado da busca em que **A** e **B** se encontrem separadas pelo menos por duas palavras e por não mais de quatro. Obviamente, cabe o uso de truncagem.

- A busca de um conceito a partir de sinônimo(s) que pode ser implementada se, a partir de estudos infométricos, são identificados os sinônimos que representam um mesmo conceito significativo e reunidos num dicionário de sinônimos. Ao buscar por uma palavra, o sistema varre a base para verificar se, entre os ‘descritores’ selecionados no processo de indexação automática, existem termos sinônimos representativos do conceito procurado, recuperando, em caso positivo, os registros de conteúdo conceitual similar, mesmo se o termo usado na busca não aparece de forma explícita. Assim, uma pesquisa por RELEVÂNCIA forneceria também registros indexados com a palavra PERTINÊNCIA. O princípio pode-se estender aos quase-sinônimos ou aos termos equivalentes em outras línguas.
- O uso de ‘*combos*’ que pode ser utilizado com vantagem para restringir a busca a um subdomínio ou uma faceta pré-determinada. Assim, um *combo* de datas permite limitar uma busca aos registros de um ano determinado, e *combos* que relacionem os tipos de documentos conservados num acervo, ou a língua em que foram redigidos, permitiriam selecionar somente os registros referentes, por exemplo, a relatórios técnicos escritos em inglês e/ou francês, e assim por diante.

#### **4. IMPLEMENTAÇÃO**

Os dados recebidos da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) e do Decanato de Pesquisa e Pós-graduação, da Universidade de Brasília, correspondentes à produção científica, técnica e artística da Universidade no

ano de 1999, foram transformados em registros de bases de dados através do programa MX, com a seguinte estrutura:

<i>Etiqueta</i>	<i>Nome do campo</i>
035	Identificação do registro
100	Autor(es)
245	Título / Fonte /Notas

O programa MX é um utilitário escrito em linguagem C, cujo propósito é trabalhar com bases de dados padrão CDS/ISIS (Unesco, 1993). Desenvolvido e distribuído gratuitamente pela Bireme – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde e membro da Organização Panamericana da Saúde (OPS) e da Organização Mundial da Saúde (OMS) –, o MX opera em todas as plataformas (MS-DOS, LINUX, UNIX, Windows, etc.) realizando a maioria das funções do software CDS/ISIS Windows (Unesco, 1997), incluindo a geração de bases de dados a partir de arquivos de texto, conversão de padrões, importação e exportação de arquivos ISO 2709 e controle de qualidade de dados.

A interface de busca do sistema aqui apresentado foi desenvolvida em WWWISIS, também denominado WXIS, utilitário igualmente distribuído pela Bireme (Bireme, 2002). Esse utilitário age como um servidor de bases de dados CDS/ISIS em um ambiente cliente/servidor Web ou Intranet. Utiliza uma linguagem de *script* baseada em XML, a IsisScript.

Os registros de uma base de dados CDS/ISIS podem ser convertidos em documentos HTML dinamicamente, ou seja, os resultados de uma pesquisa ou parte de um arquivo mestre podem ser recuperados e enviados ao cliente WWW. A saída em HTML é gerada via linguagem de formato. A possibilidade de incluir códigos HTML nas especificações do formato permite desenvolver uma interface de pesquisa

totalmente gráfica em ambiente Web ou Intranet. O WWWISIS trabalha também em várias plataformas, como LINUX e UNIX, em diferentes CPUs, e MS-DOS para Windows (95, 98, 2000) e Windows NT/XP.

O produto apresentado simula uma aplicação Web, isto é, a interface é interpretada por um servidor Web, no caso o MicroWeb, licenciado pela IndigoStar Software para a SSRR Informações (IndigoStar, 2002). O MicroWeb permite gerar um *site* Web de trabalho num CD-ROM. Usando um navegador Web (Netscape, Internet Explorer, Opera, etc.), o usuário pode executar programas cgi (*common gateway interface*), assim como visualizar arquivos cgi no CD. Para visualizar o *Website*, o usuário somente tem que executar o programa *microweb.exe*, incluso no CD, sem necessidade de qualquer configuração.

Na abertura do sistema, é apresentada ao usuário a página inicial, em cuja parte esquerda encontra-se uma janela com as seguintes opções (*Figura 2*): *Tabela do tipo de produção* (apresenta os tipos de documentos ou produções referentes às atividades científicas técnicas e artísticas, com base na codificação estabelecida pela Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior - CAPES); *Pesquisa avançada* (abre a janela para pesquisa avançada, descrita em detalhe mais adiante); *Nota explicativa* (sobre o sistema); *Créditos*; *Ajuda* (com algumas dicas básicas sobre como pesquisar); *Pesquisa rápida* (permite a consulta imediata à base, mediante a digitação do(s) termo(s) de busca e/ou de suas associações.

A página da *Pesquisa avançada* apresenta-se ao usuário como mostra a *Figura 3*. A primeira janela, *Pesquisar termos* permite a digitação das estratégias de busca, utilizando-se eventualmente de conectivos, truncagens, etc. como explicado anteriormente.

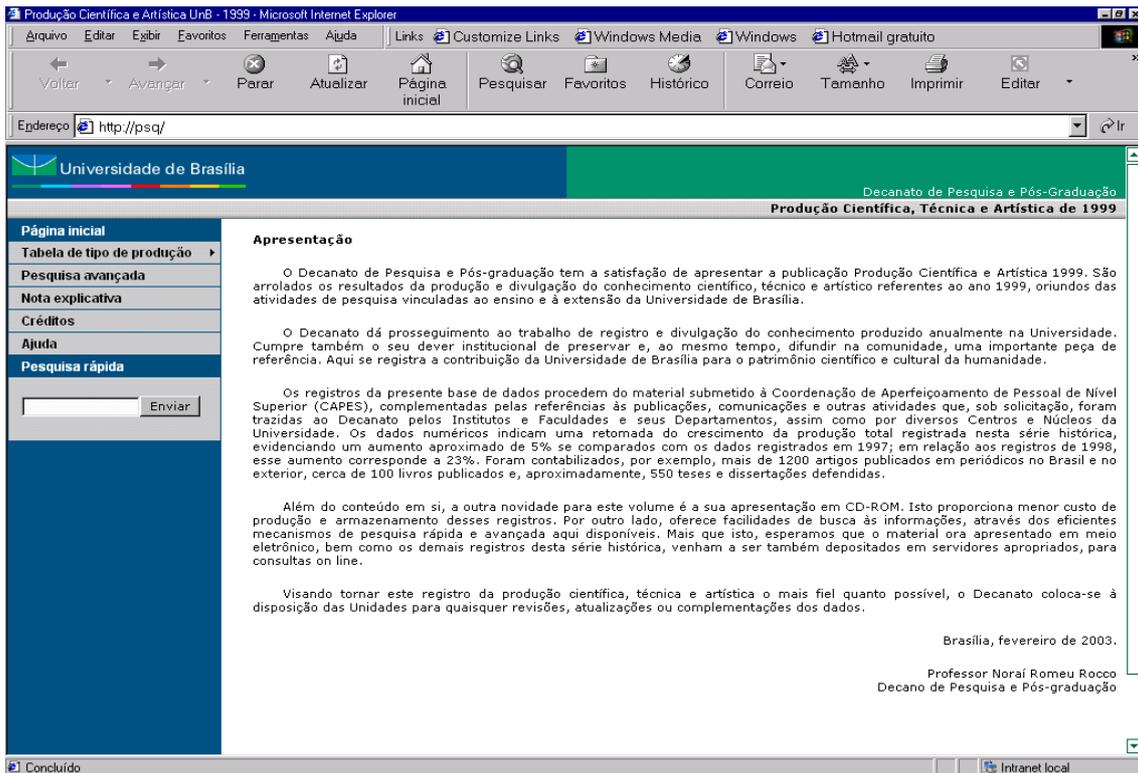


Figura 2. – Página inicial apresentada ao usuário, na abertura do sistema.

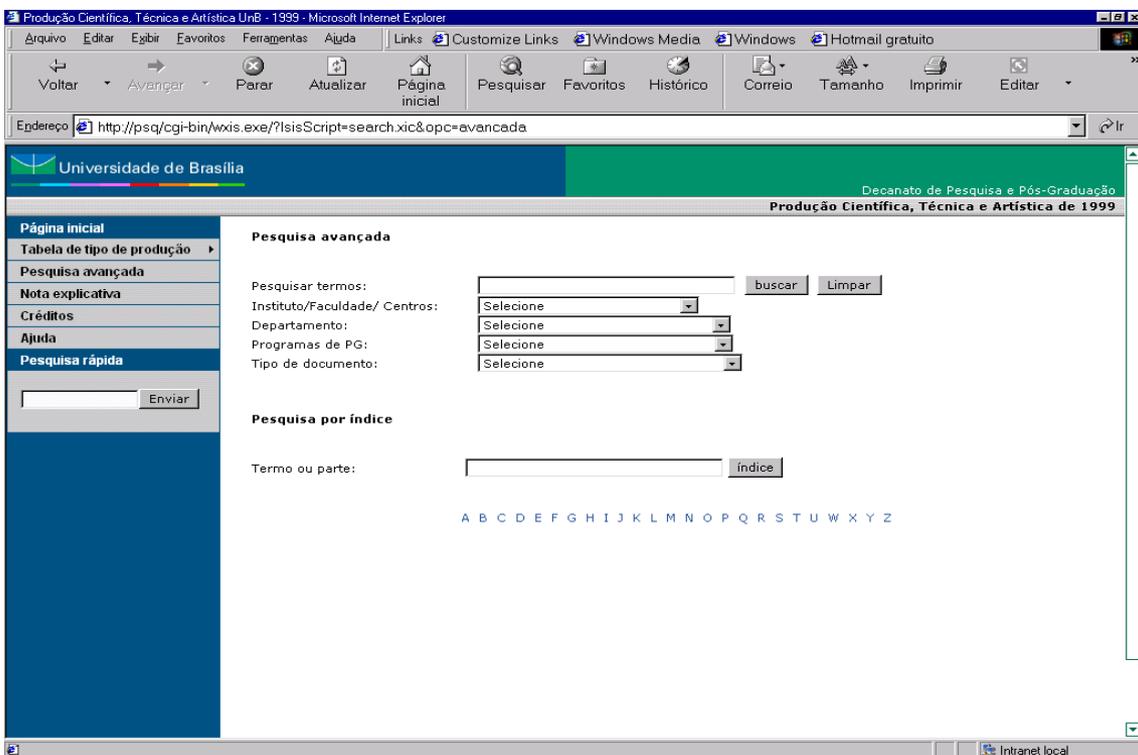


Figura 3. – Tela de Pesquisa avançada

As janelas *Faculdade/Instituto/Centro*, *Departamento*, *Programas de PG* e *Tipo de documento* são *combos* ou índices que apresentam uma lista de itens que permitem ao usuário a seleção de um item de sua escolha que pode ser associado a outro(s) itens de outro(s) *combos* e/ou a estratégia digitada na janela *Pesquisar termos*, o que permite formalizar estratégias de busca variadas, flexíveis e complexas, segundo necessidade.

A Figura 4 reproduz um fragmento do *combo* ou índice *Instituto/Faculdade/Centro* (a) e outro do *combo* ou índice *Tipo de documento* (b).

CENT AMERICA LAT E CARIBE	ARTIGO EM JORNAL BRASILEIRO
CENT DESENV SUSTENTAVEL	ARTIGO EM JORNAL ESTRANGEIRO
CENT INTERN FIS MAT COND	ARTIGOS EM REVISTA ESTRANG
FAC AGRONOMIA E MED VET	ARTIGOS EM REVISTA NACIONAL
FAC ARQUITETURA E URB	CAPITULO DE LIVRO NO BRASIL
FAC CIENCIAS DA SAUDE	CAPITULO DE LIVRO NO EXT
FAC COMUNICAÇÃO	COLETANEA NO BRASIL
FAC DIREITO	COLETANEA NO EXTERIOR
FAC EDUCAÇÃO (a)	COMPOSICOES MUSIC NO BRASIL (b)

Figura 4. – (a) Fragmento do *combo* ou índice *Instituto/Faculdade/Centro*’;

(b) Fragmento do *combo* ou índice ‘*Tipo de documento*’.

Na parte inferior da tela de *Pesquisa avançada*, apresenta-se ao usuário uma janela denominada *Pesquisa por índice* e, embaixo dela, as letras do alfabeto permitem abrir os índices de ‘descritores’, dos registros inseridos na base, que correspondem a cada letra. A *Figura 5* mostra um fragmento do índice de descritores aberto ao clicar na letra *M*. Observe-se que os ‘descritores’ – nos quais se incluem termos diversos e/ou nomes e sobrenomes de autores – vêm acompanhados de suas respectivas frequências de ocorrência, o que pode dar uma idéia da produtividade de um autor ou do peso ou da importância relativos dos termos. Observe-se também, que o índice pode ajudar a escolher a posição do símbolo de truncagem (ver, por exemplo, MADRI [português] e

MADRID [espanhol, francês, inglês...], onde a truncagem MADR\$, permite recuperar, numa busca só, o(s) registros indexados em várias línguas. O mesmo se aplica para MAFIC [inglês] e MAFICO [português], ou MAE [singular] e MAES [plural]).

#### Selecione um termo e clique em buscar

MADEIRAS [2] MADISON [1] MADRI [1] MADRID [3] MAE [1] MAES [3] MAFIC [1] MAFICO [3] MAGAGNIM [2] MAGALHAES [43]	BUSCAR
--	--------

Figura 5. - Fragmento do índice de descritores aberto ao clicar na letra M.

## 5. ALGUNS EXEMPLOS ILUSTRATIVOS

### Exemplo # 1

a) *Enunciado inicial da busca* (busca por sobrenome de autor):

MUELLER

*Resultado:* 12 registros recuperados, dentre os quais vários autores/orientadores diferentes:

**MUELLER**, BERNARDO PINHEIRO M.: 4 registros recuperados,  
**MUELLER**, CHARLES CURT: 1 registro recuperado,  
**MUELLER**, SUZANA: 4 registros recuperados,  
**MUELLER**, SUZANA PINHEIRO MACHADO: 3 registros recuperados

b) *Segundo enunciado da busca, para restringir a recuperação a um só autor / orientador:*

MUELLER and SUZ\$

*Resultado:* 7 registros recuperados, dentre os quais vários tipos de documentos (artigos publicados em revistas nacionais, artigos publicados em revistas estrangeiras, teses/dissertações orientadas...)

c) *Terceiro enunciado da busca, para restringir a recuperação a um só tipo de documento:*

MUELLER and SUZ\$ [and] 'TESES E DISSERTAÇÕES NO BRASIL' (último 'termo' escolhido no *combo* 'Tipo de Documento')

*Resultado:* 4 registros recuperados

*Índice de pertinência:* 1 (ou seja, 100 por cento).

*Índice de exaustividade:* 1 (ou seja, 100 por cento); o índice pode ser calculado porque o conteúdo da base é conhecido.

## **Exemplo # 2**

a) *Enunciado inicial da busca:*

‘INSTITUTO DE CIÊNCIAS BIOLÓGICAS’ [and] ‘ARTIGOS EM REVISTAS ESTRANG’ (expressões de busca escolhidas, respectivamente, nos *combos* ‘Faculdades/Institutos/ Centros’ e ‘Tipo de Documento’)

*Resultado:* 75 registros recuperados (distribuídos entre os diversos Departamentos)

b) *Segundo enunciado da busca, para restringir a recuperação a um só Departamento (no caso, o Departamento de Ecologia):*

‘INSTITUTO DE CIÊNCIAS BIOLÓGICAS’ [and] ‘ARTIGOS EM REVISTAS ESTRANG’ [and] ‘DEP ECOLOGIA’ (última expressão selecionada no *combo* ‘Departamento’)

*Resultado:* 36 registros recuperados (produzidos por diversos autores)

c) *Terceiro enunciado da busca, para restringir a recuperação a um só autor (no caso, Bão, Sônia Nair):*

‘INSTITUTO DE CIÊNCIAS BIOLÓGICAS’ [and] ‘ARTIGOS EM REVISTAS ESTRANG’ [and] ‘DEP ECOLOGIA’ [and] BAO (sobrenome digitado na janela ‘Pesquisar termos’).

*Resultado:* 4 registros recuperados.

*Índice de pertinência:* 1 (ou seja, 100 por cento).

*Índice de exaustividade:* 1 (ou seja, 100 por cento); o índice pode ser calculado porque o conteúdo da base é conhecido.

## **Exemplo # 3**

a) *Enunciado inicial da busca:*

CIENC\$ and inform\$

*Resultado:* 5 registros recuperados:

**Ciência da Informação:** 2 registros recuperados.

**información** [outras palavras intercaladas] **Ciencia:** 1 registro recuperado.

**Informação** [outras palavras intercaladas] **Ciênc**: 1 registro recuperado.

**ciência** [outras palavras intercaladas] **inform**ática: 1 registro recuperado.

b) *Segundo enunciado da busca, para restringir os resultados a ‘Ciência da informação’:*

CIENC\$ \$ \$ inform\$ (segundo e terceiro cifrões com um branco no meio significam uma palavra intercalada).

*Resultado:* 2 registros recuperados:

**Ciência da Informação:** 2 registros recuperados.

## **5. PROJEÇÕES FUTURAS**

Nos próximos meses, no quadro da presente parceria com o Decanato de Pesquisa e Pós-graduação da Universidade de Brasília, pretende-se:

- 1) Realizar o processamento e a geração das bases de dados correspondentes à produção científica, técnica e artística da UnB dos anos 2000 e 2001, e se possível 2002, seguindo o mesmo modelo aqui apresentado;
- 2) Reunir essas bases, numa base só, e esta com a base já existente correspondente ao período 1994 – 1998, após a necessária depuração e reindexação;
- 3) Aprimoramento da rotina (atualmente em fase de desenvolvimento e teste) de recuperação que permite o uso dos critérios de adjacência e proximidade, delimitando de forma mais flexível o número de termos intercalados entre dois termos pesquisados;
- 4) Implementar a pesquisa automática por sinônimos, e geração do dicionário de sinônimos;
- 5) Adaptação à presente aplicação da rotina anteriormente desenvolvida no quadro das pesquisas com o aplicativo InfoDoc consistente no uso de um ‘anti-dicionário’ de radicais não significativos, para aprimorar o processo de indexação automática;
- 6) Disponibilização na Internet da base de dados 1994-2002, incluindo os novos desenvolvimentos indicados.

## **6. REFERÊNCIAS BIBLIOGRÁFICAS**

BAR-ILAN, J.; PERITZ, B.C. Informetrics Theories and Methods for Exploring the Internet: An Analytical Survey of Recent Research Literature. *Libray Trends*, v.50, n.3, Winter 2002, p.372-392.

- BATES, M.J. Indexing and Access for Digital Libraries and the Internet: Human, Database and Domain Factors. *Journal of the American Society for Information Science*, v.49, 1998, p.1185-1205. (Apud HAWKINS, D.T. Information Science Abstracts: Tracking the Literature of Information Science. Part 1: Definition and Map. *Ibidem*, v.52, n.1, 2001, p.44-53.)
- BENKING, H. The need for survey- or overview information and agreement on “common frames of reference”: Information about what, by whom, under which assumption and objectives, and in which context? In: *FIS 2002: The Nature of Information: Conceptions, Misconceptions, and Paradoxes*. Electronic Conference, 6-10 May 2002. URL: [http://www.mdpi.net/-ec/ec\\_files.php?ec\\_id=1](http://www.mdpi.net/-ec/ec_files.php?ec_id=1).
- BIREME. *Interface WWWISIS, versão 5.0*. 2002. URL: <http://www.bireme.br>. (Produto licenciado pela Bireme para a SSRR Informações.)
- CHALMERS, M. Comparing Information Access Approaches. *Journal of the American Society for Information Science*, v.50, n.12, 1999, p.1108-1118.
- CLEVER PROJECT (Members), Hypersearching the Web. *Scientific American*, n.6, June 1999. URL: <http://www.sciam.com/1999/0699issue/0699raghavan.htm>.
- ELLIS, D. The Physical and cognitive paradigms in information retrieval research. *Journal of Documentation*, v.48, n.1, 1992, p.45-64
- GLÄNTZEL, W. Coauthorship Patterns and Trends in the Sciences (1980-1988): A Bibliometric Study with Implications for Database Indexing and Search Strategies. *Library Trends*, v.50, n.3, Winter 2002, p.461-473.
- GREISDORF, H. Relevance: An Interdisciplinary and Information Science Perspective. *Informing Science. – The International Journal of an Emerging Discipline* v.3, n.2, 2000, p. 67-71. URL: [informingscience.-org/Articles/Vol3/IndexV3.htm](http://informingscience.-org/Articles/Vol3/IndexV3.htm).2000.

INDIGOSTAR SOFTWARE. *MicroWeb*, 2002. URL: <http://www.indigostar.com>.

(Produto licenciado pela IndigoSTAR, Canadá, para a SSRR Informações.)

KALFATOVIC, M.R. Web-based Markup Languages: HTML, DHTML, and XML. In: ENSOR, P. (ed.). *The Cybrarian's Manual 2*. Chicago: American Library Association, 2000, pp.68-76.

LAGER, M. Spinning a Web Search,1996. URL: [www.library.ucsb.edu/untangle/lager.html](http://www.library.ucsb.edu/untangle/lager.html). Palmer, C.L. Aligning Studies of Information Seeking and Use with Domain Analysis. *Journal of the American Society for Information Science*, v.50, n.12, 1999, p.1139-1140.

LARSON, R.R. *Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace*. URL: <http://www.sherlock.berkeley.edu/asis96/asis96.html>.

O'CONNOR, B.C. *Explorations in Indexing and Abstracting: Pointing, Virtue, and Power*. Englewood, Colorado: Libraries Unlimited, 1996. 182 p. (Library and Information Science Text Series.)

ROBREDO, J.; CUNHA, M.B. DA (colab.). *Documentação de Hoje e de Amanhã : Uma Abordagem Informatizada da Biblioteconomia e dos Sistemas de Informação*. Brasília: ed. de autor, 1986. (V. também: Ed. facsimilar da 2ª ed. ampl. São Paulo: Global, 1994).

ROBREDO, J. *InfoDoc: Manual do Usuário*. Brasília DF: ed. autor, 1995. (Registrado na Library of Congress, Washington, DC sob o nº 0700479295.)

ROBREDO, J.; CUNHA, M.B. DA. Aplicação de técnicas infométricas para identificar a abrangência do léxico básico que caracteriza os processos de indexação e recuperação da informação. *Ciência da Informação*, v.27, n.1, 1998, p.11-27.

- RUGE, G., Experimentos on linguistically-based term associations. *Information Processing and Management*, v.28, n.3,1992, p.317-332.
- SOERGEL, D. The Rise of Ontologies or the Reinvention of Classification. *Journal of the American Society for Information Science*, v.50, n.12, 1999, p.1119-1120.
- SPINK, A. ; WOLFRAM, D.; JANSEN, B.J.; SARACEVIC, T. Searching the Web: The Public and their Queries. *Journal of the American Society for Information Science*, v.52, n.3, 2001, p.226-234.
- UNESCO, *Mini-micro CDS/ISIS Reference Manual* (versão 2.3-3.071). Paris: Unesco, 1993;
- UNESCO, *CDS/ISIS Windows version*. 1997. URL: <ftp://ftp.unesco.org/pub/winisis/>
- WILSON, T.D. Human information behavior. *Informing Science. – The International Journal of an Emerging Discipline* v.3, n.2, 2000. URL: [informingscience.org/Articles/Vol3/IndexV3.htm](http://informingscience.org/Articles/Vol3/IndexV3.htm).
- WOLFRAM, D. Term Co-ocurrence in Internet Search Engine Queries: An Analysis of the Excite Data Set. *Canadian Journal of Information and Libray Science*, v.24, n.2/3, 1999, p.12-33.
- WOLFRAM, D. Applications of Informetrics to Information Retrieval Research. *Informing Science – The International Journal of an Emerging Discipline*, v.3, n.2, 2000, p.77-82. URL: [informingscience.org/Articles/Vol3/IndexV3.htm](http://informingscience.org/Articles/Vol3/IndexV3.htm).