



ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS

ENSAIOS APB

**NAVEGAR É PRECISO:  
como entender a estrutura  
de busca na Web**

**Patrícia Zeni Marchiori**

**Ensaio APB, n. 80**

*APB - ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS - APB*

**NAVEGAR É PRECISO:  
como entender a estrutura  
de busca na Web**

**Patrícia Zeni Marchiori**

**Ensaio APB, n. 80**

**APB - ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS - APB**

**NAVEGAR É PRECISO:  
como entender a estrutura de busca na Web**

**Patrícia Zeni Marchiori**

**Ensaio APB, n. 80**

**São Paulo  
Julho  
2000**

- 1 - MELO, José Marques de. Comunicação de Massa x Leitura. 1994.
- 2 - MOSTAFA, Solange Puntel. Balcão de Informações: o mercado emergente. 1994.
- 3 - TAVARES, Maria Christina de Moraes. Atuação da Biblioteca Infanto-Juvenil. 1994.
- 4 - MURGIA, Eduardo. A Crise da Informação. 1994.
- 5 - OLIVEIRA, Silas Marques de. A Crise dos recursos Humanos em Bibliotecas. 1994.
- 6 - BARROS, Maria Helena T. C. de. A Atuação da Biblioteca Escolar: relato de uma crise. 1994.
- 7 - DIAS, Maria Cristina Santarém et al. Alternativas para Contornar a Crise da Leitura: uma experiência do ônibus-biblioteca na cidade de São Paulo. 1994.
- 8 - FERREIRA, Marta José et al. Projeto "Soma". 1994.
- 9 - LARROUDE, Rita Luisa et al. Terceira Idade: relato de uma experiência, 1991-1992. 1994.
- 10 - SILVA, Helen de Castro et al. Um espaço para a Fantasia. 1994.
- 11 - TOMAZELLI, Angela M. et al. Criança de Periferia não Lê: desmistificação. 1994.
- 12 - RIVA, Eliane Barbosa et al. Terceira Idade: programa integrado. 1994.
- 13 - ALMEIDA JÚNIOR, Oswaldo Francisco de. O Espaço da Biblioteca: uma reflexão. 1994.
- 14 - VALENTIM, Marta Lígia Pomim. Leitura Técnica e seu Papel na Pesquisa & Desenvolvimento. Jan. 95.
- 15 - ALMEIDA JÚNIOR, Oswaldo Francisco de. Biblioteca pública: ambigüidade, conformismo e ação guerrilheira do bibliotecário. Fev. 95.
- 16 - VALLS, Valéria. O espaço do bibliotecário no gerenciamento de documentos do Sistema da Qualidade. Mar. 95.
- 17 - CARDIN, Tânia Maria Sanvezo. Lixo reciclável x incentivo à leitura: uma relação que deu certo no município de Ipirorã - PR. Abr. 95.
- 18 - LIMA, Justino Alves. Bibliotecas e bibliotecários: o perfil de um caso. Maio 95.
- 19 - MODESTO, Fernando. Apontamentos sobre a ergonomia na implantação e uso do computador na biblioteca. Jun. 95.
- 20 - CÔRTE, Adelaide Ramos e. Memória técnica. Jul. 95.
- 21 - FUJINO, Asa. A gestão da informação no processo de cooperação universidade-empresa: uma visão crítica. Ago. 95.
- 22 - FÁRIA, Ivete Pieruccini. Livro e leitura no Brasil: alguns aspectos acerca da entrada do impresso no país. Set. 95.
- 23 - SMIT, Johanna. Algumas questões sobre os documentos audiovisuais em bibliotecas. Out. 95.
- 24 - SILVA, A. M. S., ALMEIDA, G. M. A. B., BELLUZZO, R. C. B. O Plano de Gestão da Qualidade e sua implantação na rede de bibliotecas da UNESP: relato de uma experiência. Nov. 95.
- 25 - VERGUEIRO, Waldomiro C. S. Gestão da Qualidade e Bibliotecas Públicas: o difícil caminho para as instituições brasileiras. Dez. 95.
- 26 - LANE, Sandra S., VAL, Marta R. S. Ribeiro do. Preservação de acervos de bibliotecas: Parte I. Degradação dos materiais. Jan. 96.
- 27 - LANE, Sandra S., VAL, Marta R. S. Ribeiro do. Preservação de acervos de bibliotecas: Parte II. Um modelo de programa local. Fev. 96.
- 28 - SOUZA, Marta Alves de. Internet: a rede global. Mar. 96.
- 29 - MODESTO, Fernando. Combate ao vírus de computador na biblioteca. Abr. 96.
- 30 - BARTALO, Linete et al. A importância da leitura na formação do professor. Maio. 96.
- 31 - ARAÚJO, Eliany Alvarenga de. Sociedade de informação: espaço da palavra onde o silêncio mora? Jun. 96.
- 32 - GUIMARÃES, José Augusto Chaves. A Legislação profissional do bibliotecário. Jul. 96.
- 33 - MARTUCCI, Elisabeth Márcia. Abordagem qualitativa de pesquisa em biblioteconomia: uma introdução. Ago. 96.
- 34 - MARCHIORI, Patrícia Zeni. Eram os deuses astronautas? ou São os bibliotecários, profissionais da informação? Set. 96.
- 35 - FERREIRA, Sueli Mara S. P., KROEFF, Márcia S. Referências bibliográficas de documentos eletrônicos: vol. 1. Out. 96.
- 36 - FERREIRA, Sueli Mara S. P., KROEFF, Márcia S. Referências bibliográficas de documentos eletrônicos: vol. 2. Nov. 96.
- 37 - ALMEIDA JÚNIOR, Oswaldo Francisco de. Roubo, depredação de materiais e campanhas educativas em bibliotecas: proposta de um modelo de avaliação. Dez. 96.
- 38 - SOUZA, Francisco das Chagas de. O bibliotecário brasileiro e seu humanismo. Jan. 97.
- 39 - LIMA, Justino Alves. Mobilização para uma política de conservação e manutenção de acervos contra o agente biológico humano. Fev. 97.
- 40 - SMIT, Johanna W., MACAMBYRA, Marina M. Tratamento de multimídia. Mar. 97.
- 41 - SANTOS, Jussara Pereira. O ensino de biblioteconomia no Mercosul: propostas de integração e harmonização curricular. Abr. 97.
- 42 - FUJITA, Mariângela Spotti Lopes. Elaboração de tesaurus monolíngues com o programa TECER: considerações sobre o uso. Maio 97.
- 43 - BARRETO, Angela Maria. Conversas com quem gosta de informar. Jun. 97.
- 44 - LIMA, Justino Alves. As entidades da biblioteconomia: uma tentativa de globalização e uma iniciativa de intervenção política. Jul. 97.
- 45 - TALAMO, Maria de Fátima G. M. Linguagem documentária. Ago. 97.
- 46 - MODESTO, Fernando. O bibliotecário e o mercado de trabalho: alguns comentários. Set. 97.
- 47 - RECINE, Analúcia Viviani dos Santos. Análise de partituras. Out. 97.
- 48 - TOMAEL, Maria Inês. Informação e globalização: reflexos de uma nova era. Nov. 97.
- 49 - FIGUEIREDO, Nice. Repensando a biblioteca pública brasileira: considerações em torno de resultados de pesquisa. Dez. 97.
- 50 - FIGUEIREDO, Nice. Repensando a biblioteca universitária brasileira: como prosseguir - notas para um projeto de pesquisa. Jan. 98.
- 51 - FERREIRA, Margarida M. Estudo do formato para registro bibliográfico Marc: volume 1. Fev. 98.
- 52 - FERREIRA, Margarida M. Estudo do formato para registro bibliográfico Marc: volume 2. Mar. 98.
- 53 - RUSSO, Mariza. Bibliotecas universitárias brasileiras: diretrizes para o próximo milênio. Abr. 98.
- 54 - GAUZ, Valéria. O bibliófilo José Mindlin: impressões de um encontro. Maio. 98.
- 55 - COSTA, Márcia Betânia da. Implantação do movimento 5S's em unidades de informação. Jun. 98.
- 56 - SIQUEIRA, Maria das Graças. Ler é conquistar autonomia. Jul. 98.
- 57 - SOUZA, Marta Alves de. Fontes de informação em Ciências Exatas: uma síntese. Ago. 98.
- 58 - FIGUEIREDO, Nice. A automação das bibliotecas universitárias: resultado de pesquisa. Set. 98.
- 59 - MODESTO, Fernando. O bibliotecário e o mercado de trabalho: estratégias para o emprego. Out. 98.
- 60 - BROWN, Doris R. O consórcio nas bibliotecas acadêmicas dos EUA. Nov. 98.
- 61 - GOMEZ, Margarita Victoria. Educação e informática: caminho entrelaçado com a biblioteconomia. Dez. 98.
- 62 - LIMA, Vânia Mara Alves. Comunicação e representação documentária. Jan. 99.
- 63 - BLATTMANN, Ursula, DUTRA, Sigrid Karin Weiss. Atividades em bibliotecas colaborando com a educação a distância. Fev. 99.
- 64 - FIGUEIREDO, Nice. Automação das bibliotecas universitárias: a visão dos usuários. Mar. 99.
- 65 - ALMEIDA, Elisângela Lino de. Conservação e acondicionamento de discos de vinil e fitas cassetes. Abr. 99.
- 66 - OLIVEIRA, Ana Lúcia Antunes de. A biblioteca escolar no processo de ensino-aprendizagem: uma experiência. Maio. 99.
- 67 - ALMEIDA JÚNIOR, Oswaldo Francisco de. Agora informacional. Jun. 99.
- 68 - FREIRE, Bernardina M. Juvenal, PEREIRA, Raquel G., LIMA, Geysa F. C. de. Biblioteca volante em canteiro de obras: relato de uma experiência. Jul. 99.
- 69 - FIERLI, Aglaê de Lima, CATARINO, Maria Elisabete. Classificação Decimal de Dewey em CD-ROM. Ago. 99.
- 70 - MARQUES, Eliana Maria. Biblioteca Pública no Brasil: sonho ou realidade? Set. 99.
- 71 - VALENTIM, Marta Lígia Pomim. A atividade de investigação em Ciência da Informação. Out. 99.
- 72 - PEREIRA, Enidêli A. Zaquia et al. Agentes de Tecnologia: uma experiência de estágio na área de informação e gerência do Curso de Biblioteconomia da UEL. Nov. 99.
- 73 - SOUZA, Samuel R. M. de. Como fazemos as coisas por aqui? Bibliotecários e Cultura Organizacional. Dez. 99.
- 74 - VERGUEIRO, Waldomiro. Qualidade em serviços de informação: o foco no cliente. vol. 1. Jan. 2000.
- 75 - VERGUEIRO, Waldomiro. Qualidade em serviços de informação: o foco no cliente. vol. 2. Fev. 2000.
- 76 - DUTRA, Miriam Regiane. A indústria da informação no Brasil: reflexões. Mar. 2000.
- 77 - TOMAEL, Maria Inês et al. Critérios para avaliar fontes de informação na Internet. Abr. 2000.
- 78 - TOMAEL, Maria Inês et al. Fontes de informação na Internet: acesso e avaliação das disponíveis nos sites das Universidades. Maio 2000.
- 79 - MOLOGNI, Micherle. Programa INFOINDEX: a agilidade no trabalho de classificação e indexação. Jun. 2000.
- 80 - MARCHIORI, Patrícia Zeni. Navegar é preciso: como entender a estrutura de busca na Web. Jul. 2000.

## NAVEGAR É PRECISO como entender a estrutura de busca na Web

Patricia Zeni Marchiori (<sup>1</sup>)

- O que são sistemas de busca na Web?

A Internet contém informações valiosas em meio ao que se poderia chamar de "caos informativo". Na verdade, a questão de recuperação da informação neste ambiente interconectado, livre de censuras e fronteiras, repousa em tripla responsabilidade: Primeiramente, daqueles que disponibilizam conteúdos, logo daqueles que se dispuseram a "organizá-lo" e, inevitavelmente, daqueles que buscam informações, quaisquer que sejam seus motivos.

Embora não se possa determinar quantas "páginas" estão disponíveis na *web* (estatísticas variam de 1 bilhão a 5 bilhões - com uma taxa de crescimento de 1/3 a cada ano), isto não diminui a ansiedade em saber que a Internet é, definitivamente, um repositório imenso e crescente de dados e informação. Sob o ponto de vista de um profissional da informação, este volume de informação potencial, tem desafiado os esforços de indexação e classificação de conteúdos, que podem auxiliar na efetiva recuperação da informação por parte dos internautas.

Sob esta perspectiva, a parte gráfica da Internet (a *web*), tem sido classificada e reclassificada das mais diferentes maneiras. Pode-se falar que, em seus quase dez anos de existência, a busca por informações na *web* já passou por 4 fases que convivem de maneira harmoniosa e interdependente, tal é a complexidade da tarefa e as diferentes demandas de usuários. A primeira é representada pelas ferramentas de busca e dos diretórios de *sites* (Altavista, Yahoo!, Google, por exemplo), a segunda, pelas meta-ferramentas de busca (a "família Miner", Copérnico, Metacrawler), a terceira, os portais (e vortais) – voltados para compilação de conteúdos e *links* significativos em áreas temáticas determinadas – e, mais recentemente, os chamados de "serviços personalizados", em que o usuário escolhe o tipo de informação que quer acessar e adapta o visual de suas páginas. Embora não se possa dizer qual o futuro da recuperação da informação disponível na Internet, o objetivo do usuário será sempre buscar uma informação "de qualidade". Sob este ponto de vista, quanto "melhor" o usuário souber "indagar" o sistema de busca de sua escolha, maior é a possibilidade de encontrar o que deseja.

---

<sup>1</sup> Doutora pela ECA/USP. Prof. Adjunto I, Departamento de Ciência e Gestão da Informação, Universidade Federal do Paraná. e-mail [pzeni@coruja.humanas.ufpr.br](mailto:pzeni@coruja.humanas.ufpr.br)



### LEMBRE-SE:

1. Sites e páginas web devem ser vistas sob dois pontos de vista  
**do webdesigner** (autor da página/site) – que vai elaborar as páginas para serem indexadas pelos sistemas de busca, da maneira mais fácil e rápida possível  
**do usuário/pesquisador** – que quer encontrar os sites/páginas que apresentem um determinado assunto ou abordem aspectos específicos de um assunto
2. Os sistemas de busca usam diferentes estratégias, e têm diferentes interfaces – como se fossem diferentes bibliotecas com diferentes bibliotecários

## • Qual a responsabilidade do pesquisador ?

O verbo “garimpar” se aplica bem à pesquisa de informação na Internet. O resultado final de uma busca vai depender muito de como ela começa. Isto implica na maior ou menor habilidade do internauta em :

- Definir a(s) estratégia(s) de busca;
- Escolher o sistema de busca que, potencialmente, pode ser mais efetivo;
- Dominar as estratégias de busca que cada sistema apresenta.

## • O que fazer primeiro?

- 1) **Aprender a analisar o tópico/tema a ser pesquisado**
  - para que você vai usar a informação?
  - que você já “sabe” sobre esta informação?
- 2) **Identificar os principais conceitos da pesquisa**
- 3) **Escolher/definir as palavras-chave que descrevam os conceitos**
- 4) **Determinar se existem sinônimos, termos relacionados ou outras variações para as palavras-chave que foram definidas**

quanto mais específica for uma palavra/frase, mais eficiente será a busca:

A busca por - cercopithecus aethiops – pode dar melhores resultados que - african green monkey - em especial quando esta informação vai ter um uso científico/acadêmico

cuidado com a ortografia

digitar astronut vai acarretar resultados diferentes que astronaut

usar variações pertinentes

fly-fishing = flyfishing = fly fishing

quando possível, considere a utilização do operador booleano “NOT” (ou o equivalente disponível no sistema)

uma busca com “metal NOT heavy NOT music “- vai trazer sites que tratam de metais industriais e não de bandas de rock pesado

verificar que outros aspectos devem/podem ser definidos no sistema de busca: truncagem de termos, operadores de proximidade, operadores booleanos não usuais, por exemplo. Neste particular, tais "ajustes" na pesquisa vão depender do sistema utilizado

- 5) escolher o(s) sistema(s) de busca
- 6) ler as instruções e arquivos de ajuda dos sistemas de busca
- 7) criar uma "expressão de busca"  
explorar as opções e a sintaxe apropriada para o sistema escolhido
- 8) refinar a pesquisa (se necessário – e normalmente, é!)
- 9) avaliar os resultados  
os resultados obtidos são relevantes quando comparados com as perguntas iniciais?
- 10) REPETIR os passos 6,7, 8 e 9 em outro(s) sistema(s) de busca



#### LEMBRE-SE:

##### ■ pesquise com "visão periférica":

A cada grupo de resultados que você recuperar, **não desanime com o total de itens** – examine, rapidamente, pelo menos os vinte primeiros resultados. Nesta breve leitura dinâmica você pode encontrar termos que refinem a pesquisa (termos que podem/devem ser excluídos, termos que são importantes sinônimos ou termos que devem ser incluídos e que "escaparam").

## • Que é um "mecanismo de busca"?

É um tipo de sistema de busca que cria seus "índices" de forma automática (sem a intervenção de seres humanos), utilizando *software* apropriado, chamados de *spiders* ou *crawlers*. Um "robô" (*spider*) vasculha constantemente os servidores *web* (também podendo fazê-lo nos grupos de discussão, nos servidores FTP e *Gopher*) e constrói um índice de todo o material encontrado. Nesta varredura, o robô pode encontrar novos *sites*, atualizar os que já compõem a base de dados do sistema, "deletar" as URLs que foram retiradas do servidor (o famoso erro 404). Alguns destes sistemas podem, inclusive, "clonar" as páginas e não somente remeter o internauta ao *site* original

De maneira geral, todo robô busca indexar em cada página, a URL ela mesma, o título da página, os cabeçalhos que abrem cada seção e outras *tags* e/ou *metatags* de hipertexto. Outros, ainda, registram as palavras mencionadas com mais frequência ou as primeiras linhas do texto, que indica sua estrutura de recuperação de informação (baseada, também em geral, na frequência da palavra no texto, ou seu aparecimento nas *metatags* mais "relevantes" - como a de título, por exemplo). Estas peculiaridades vão definir a profundidade de busca da base de dados do sistema de busca, assim como suas características de revocação e precisão.

Igualmente, os sistemas de busca retornam periodicamente às páginas indexadas para verificar mudanças – esta periodicidade varia de mecanismo para mecanismo e influi diretamente na qualidade e atualidade da base de dados. Assim, o criador da página ou *site*, não necessita obrigatoriamente solicitar a inclusão de seu material na base de dados de determinado mecanismo. Uma hora ou outra, a *spider* do mecanismo vai encontrar o recurso.

A maioria dos instrumentos de busca fornece um conjunto próprio de operadores, delimitadores e regras que auxiliam nas buscas

## EXEMPLOS DE SISTEMAS BASEADOS EM ROBÔS

Cobertura nacional	URL
Bookmarks	<a href="http://www.bookmarks.com.br">www.bookmarks.com.br</a>
Radaruol	<a href="http://www.radaruol.com.br">www.radaruol.com.br</a>
TodoBr	<a href="http://www.todobr.com.br">www.todobr.com.br</a>
Radix	<a href="http://www.radix.com.br">www.radix.com.br</a>
Internacionais	URL
Altavista	<a href="http://www.altavista.com">www.altavista.com</a>
HotBot	<a href="http://www.hotbot.com">www.hotbot.com</a>
Excite	<a href="http://www.excite.com">www.excite.com</a>
Lycos	<a href="http://www.lycos.com">www.lycos.com</a>
Northern Light	<a href="http://www.northernlight.com">www.northernlight.com</a>

### • Que é um “diretório” ou “catálogo”?

É um tipo de sistema de busca que, diferentemente dos sistemas com base em robôs, tem o conteúdo de sua base de dados criado por seres humanos. Os *sites* devem ser submetidos manualmente (pelo interessado) ou esperar que um “editor” deste catálogo tenha notícia ou “esbarre” com o *site*, para que este seja catalogado. Quando são submetidos ou propositadamente inseridos na base de dados, são alocados em um sistema de hierarquias/categorias, o que facilita a busca para os internautas que têm uma certa idéia do que buscam, mas que não têm muita familiaridade com os recursos de lógica booleana, por exemplo. O exemplo internacional mais conhecido é o Yahoo! – no Brasil, pode-se citar o Cade?

### • Que outros tipos de sistemas de busca existem?

- Os “diretórios anotados”, que apresentam sumários dos *links* das páginas, que pretendem ajudar o usuário a verificar o que o *site* apresenta, antes de se aventurar a navegar no referido *link*. Diferentemente dos robôs e dos catálogos (que “juntam” links em suas bases de dados), em tais diretórios, os *abstracts* são feitos por pessoas e o usuário pode ler um pouco sobre o *site* - antes de ir até ele - e decidir se vale a pena consultá-lo ou não.  
Exemplos: eBlast; Lycos Top 5%; Magellan
- Os “diretórios especializados”, que abrangem “tipos” de dados, tais como: endereços de *e-mail* (<http://www.whowhere.com>), informações sobre servidores *web*,

listas de listas de discussão (<http://www.liszt.com>) , servidores que oferecem free/shareware (<http://www.tucows.com>), por exemplo.



#### CURIOSIDADE:

*Por utilizarem um software (e não seres humanos), os sistemas de busca com base em robôs, possuem bases de dados quantitativamente maiores, o que não significa que a qualidade de recuperação acompanha o tamanho da base de dados disponível.*



#### ATENÇÃO !

*A nova tendência para os sistemas de busca está na estrutura desenvolvida por dois estudantes da Universidade de Stanford (EUA). O sistema inventado por eles (<http://www.google.com>), utiliza uma tecnologia que relaciona os resultados da pesquisa por ordem de importância. Isto ocorre porque a base de dados é construída sob a premissa de que o site que aparece em primeiro lugar nos resultados, é aquele que recebeu o maior número de links de outros sites. Como cada link é tido como uma "indicação", quanto mais o site é indicado (citado), mais importante ele é, ou, pelo menos, indica que o assunto de que trata é "quente". Esta tendência se fortaleceu a ponto do Altavista ter criado um sistema similar (o Raging) e o Yahoo! ter adotado tal tecnologia.*

- Qual o uso da lógica booleana nos sistemas de busca?

Quando se busca uma informação, expressa-se o tema de interesse por meio de várias palavras ou expressões combinadas, como por exemplo: "quais as reações secundárias da efedrina?"; "como os computadores são utilizados no ensino?"; "como construir com elementos pré-fabricados?"; "quais as técnicas para fabricar alfinetes?"

Instintivamente, sabemos que somente frente a um conjunto muito reduzido de itens se poderá pensar em localizar TUDO que existe sobre um determinado tema. Todavia, em bases de dados que têm milhões de referências (e que tendem a ter ainda mais), a necessidade de precisão na busca é crucial.

Uma boa parcela dos sistemas de busca oferece a opção de recuperação por "operadores lógicos", também chamados de "operadores booleanos". Considerando-se que há uma relação necessária entre os conceitos alfinete/fabricação, efedrina/reações (utilizando-se os exemplos acima), esta pode ser estabelecida pelos operadores booleanos no momento da busca. Uma vez que as máquinas não podem pensar, é necessário "traduzir" a questão, ajudando o computador a "entender" a relação entre os conceitos que se deseja recuperar.

Os operadores são de uso universal e podem ser simbolizados de diferentes maneiras. De acordo com a relação estabelecida (normalmente a restrição ou a expansão) entre vários termos, o significado da pergunta varia e também varia a informação obtida como resposta

De maneira geral, na pesquisa por operadores booleanos, se utilizam os conectores:

AND (E), OR (OU), NOT (NÃO)



### ATENÇÃO !

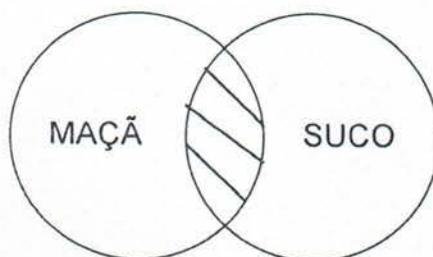
Os operadores (conectores) booleanos devem ser digitados em maiúsculas. Caso você digite em minúsculas, muitos sistemas de busca vão considerar o conector como uma palavra-chave, ou mesmo como uma palavra não significativa (stop word) e não um operador lógico. Isto comprometerá os resultados.

Observação: isto não é uma regra para todos os sistemas de busca, mas, na dúvida, sempre digite os conectores em maiúsculas e os termos/descriptores em minúsculas, pois muitos instrumentos são *case sensitive*.

## O operador “E”

Quando se buscam informações sobre suco de maçã, os itens recuperados devem necessariamente conter o termo **maçã** e o termo **suco** ao mesmo tempo.

Representação em gráficos de Venn:



Outros exemplos:

- java AND microsoft
- the chamber AND movie

## O operador “OU”

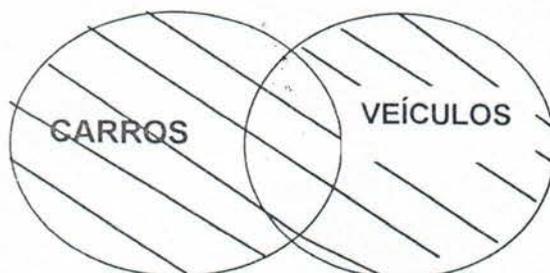
Apesar do desenvolvimento de *softwares* que trabalham com funções semânticas, o computador não é, ainda, um ser pensante. Termos com o mesmo significado, tais como EDIFÍCIO, HABITAÇÃO, MORADIA., não são automaticamente inferidos pelo computador, e isto se reflete na estruturação de pesquisas na Internet. Alguns sistemas trabalham com a chamada *fuzzy logic*, que permite que a base de dados seja indagada em termos de conceitos relacionados em função do termo/descriptor inserido na caixa de pesquisa.

Assim, para se aumentar as possibilidades de recuperar a maior quantidade de informação, é conveniente usar sinônimos (termos relacionados ou similares).

Por ex.: **planta OU erva OU vegetal**

O conectivo OR (OU) conecta dois termos e expande a pesquisa. Por exemplo, se você pesquisar por "carros ou veículos" os itens que tiverem quaisquer um destes termos serão recuperados.

Representação em gráficos de Venn:



Os termos não precisam ser necessariamente sinônimos, como por exemplo, quando se busca por **biogás OU metano**. Neste caso, qualquer um dos termos pode ser utilizado para recuperar informação sobre combustível obtido por fermentação anaeróbia de dejetos orgânicos.

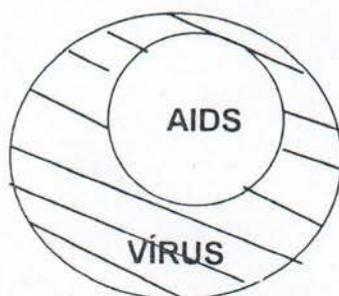
Assim, se toma o conjunto de itens com o termo **biogás** e o une ao conjunto de itens com o termo **metano**. Forma-se assim, um terceiro conjunto (com a união de ambos) que conterá documentos com um ou outro ou ambos os termos

O operador OU também é útil para pesquisas quando não se tem certeza sobre a maneira de se escrever o descritor, como por exemplo: **cusco OU cuzco**, ou mesmo para se expandir a pesquisa quando se buscam sinônimos (ex: menino OU piá OU garoto).

## O operador "NÃO"

O conectivo NOT (NÃO) restringe a pesquisa, pela exclusão de itens que contêm o segundo termo de pesquisa. Por exemplo, se você busca por **vírus e NÃO AIDS** - todos os itens contendo "vírus" vão ser recuperados - à exceção daqueles que contém a palavra "AIDS"

Representação em gráficos de Venn:



## O operador "NEAR"

O operador NEAR indica que o termo (ou a palavra-chave) de pesquisa obrigatoriamente deve aparecer entre "x" palavras de outro termo ou palavra-chave. Os sistemas de busca que permitem tal operador, definem tal número de palavras. No Altavista, por exemplo, os termos se separam por 10 outras palavras/termos, enquanto que no Open Text, estão separadas por 80 palavras.

Exemplo:

vulcão NEAR lava



### ATENÇÃO !

#### Recurso de *truncagem* e *stemming*

Alguns sistemas de busca permitem a busca pela "raiz" de uma palavra – o que permite a recuperação por distintas variações no final desta. Em alguns instrumentos isto é feito automaticamente, em outros, como no Altavista, por exemplo, deve-se utilizar o asterisco (\*)

Por exemplo:

auto\* (permite recuperar, "automático", "automóvel", "automóveis")

**MAS CUIDADO !**

Car\* = carro; careta; carimbo

## • O que é a "matemática" dos sistemas de busca?

Alguns sistemas de busca permitem o uso de uma "matemática" de busca, na qual os operadores booleanos são representados por símbolos de adição e subtração, que têm a mesma função dos booleanos, ou seja, ampliar ou restringir a pesquisa.

A possibilidade de uso destes sinais, permite uma composição de busca mais detalhada, na qual se podem "multiplicar" termos e mesmo indicar, como em uma equação, que termos o instrumento deve buscar primeiro e que termos são incluídos depois no processo.

## O símbolo de "+"

Como para o booleano AND, este sinal é utilizado quando se deseja que o instrumento de pesquisa encontre páginas que tenham todas as palavras que foram solicitadas e não apenas algumas delas.

Exemplos:

- Uma página que tenha referência tanto ao Presidente Cardoso como ao Senador Magalhães, pode ser digitada como:

+Cardoso +Magalhães

Desta forma, somente páginas que tenham ambas as palavras vão aparecer nos resultados.

Outros exemplos:

+windows +98 +falhas  
(evita páginas falando do *windows* em geral)

+guerra +estrelas +ameaça +fantasma  
(recupera páginas sobre os filmes da série Guerra nas Estrelas, mas que falem só sobre o episódio Ameaça Fantasma)

Este símbolo é muito útil quando a pesquisa inicial recuperou um número muito grande de informações (*hits*)

Exemplos:

- Imagine que o internauta quer reservar um espaço em um *camping* no Quartelá. É possível iniciar a busca simplesmente digitando

quartelá

Porém, há grandes chances dele ser afogado por resultados. Em vez disto, o mais adequado é pesquisar por todas as palavras que o internauta conhece, e que podem aparecer na página que procura. Uma estratégia de busca mais refinada, seria:

+quartelá +camping +reservas



#### ATENÇÃO !

*NÃO* deixe espaços entre o sinal e o termo. Porém, deixe um espaço entre o descritor anterior e o sinal gráfico que vem antes do próximo descritor.

## O símbolo de “-“

Corresponde ao booleano NOT, e deve ser usado (se o instrumento assim o permite) para encontrar páginas que apresentem uma palavra, mas que não tenham outra (ou outras), como por exemplo:

- Imagine que o pesquisador quer uma informação sobre o presidente Clinton mas NÃO quer ser afogado por páginas que falam do seu escândalo com Monica Lewinsky. Neste caso, uma expressão “matemática” adequada, seria:

+clinton -lewinsky

Isto “diz” ao instrumento de pesquisa que busque por páginas que mencionem “clinton” mas remova aquelas que mencionam também “lewinsky”

## Usando aspas para multiplicar

Muitos sistemas de busca (não importa se utilizam booleanos ou *power search* ou ainda a "matemática de pesquisa", permitem que se "multipliquem" os termos de pesquisa, que passam a dar origem a uma "frase de pesquisa" – que auxilia a especificar ainda mais a expressão de busca.

Por exemplo:

Quando da busca por páginas que falavam sobre reservas no camping do Guartelá, o internauta pode digitar:

+guartelá +camping +reservas

Esta expressão recupera páginas que têm todas as palavras nela, mas isso não garante que essas palavras estejam, necessariamente, perto uma das outras.

Uma página pode mencionar "guartelá" no parágrafo inicial e depois falar sobre reservas de *camping* na Ilha do Mel. De certa maneira, todas as palavras estão na página, mas ainda assim, podem não representar o que se deseja.

Compor uma frase de pesquisa evita este problema. O pesquisador "diz" ao instrumento de pesquisa que retorne páginas onde os termos aparecem na ordem exata que foi especificada. No exemplo acima, basta colocar aspas para formar a frase:

"reserva para *camping* no Guartelá"

Assim, só páginas que têm todas as palavras na exata ordem acima serão listadas. As respostas deverão ser muito mais pertinentes que apenas usando o sinal de adição.

Contudo, a pesquisa por frases pode limitar demasiadamente a busca, uma vez que o sistema se limitará a comparar a frase com os índices que possui em sua base de dados. Caso a frase for mal formulada, ou mesmo tiver plurais ou variações de palavras que não constam dos índices, o retorno será nulo ou inconsistente, mesmo que a base contenha efetivamente, material sobre o tema desejado.



### COMBINAÇÃO DE SÍMBOLOS

Por exemplo, ao se digitar:

+campeonato +tenis -wimbledon -"Roland Garros"

esta expressão "matemática" recuperará os campeonatos de tênis que não sejam nem de Wimbledon e nem de Roland Garros

## • O que é "Power Search"?

Além da "matemática" - e antes de se trabalhar com a lógica booleana - alguns sistemas de busca, tratam de facilitar ao máximo a tarefa do internauta. Ao invés de solicitar que este componha sua busca, o sistema já apresenta opções "clicáveis" que representam restrições, ampliações e outras limitações à busca. As principais estruturas de *power search* indicam os operadores booleanos mais comuns e a pesquisa por frases. Além disto, o pesquisador pode encontrar outras possibilidades de restringir sua busca, por data, língua, domínio, entre outros recursos.

A maior parte dos sistemas de busca que oferecem a opção de busca via *power search* trabalha com menus ativados por caixas de comando (combos).

## *Match any* = qualquer das palavras

Comparável ao booleano AND (e ao sinal de "+") é utilizado para pesquisas que incluam "qualquer um dos termos da expressão de busca".

Algumas ferramentas farão isto por "*default*", tais como o Altavista, o Yahoo!, o Infoseek e o WebCrawler. Outros sistemas solicitaram que se indique o tipo de pesquisa na "caixa de pesquisa" ou em uma página chamada "*advanced search*" (pesquisa avançada), tais como o HotBot e o Lycos

## *Match all* = todas as palavras

Comparável ao booleano OR, permite que se busquem páginas que contenham todos os termos de pesquisa utilizados - e não qualquer um deles.

## *Exclude* = excluir

Comparável ao booleano NOT, indica para o sistema de busca que se deseja excluir resultados que contenham a(s) palavra(s) determinada(s)

## *Phrase Searching* = frase inteira

Novamente, é a utilização de termos em uma ordem e grafia definida pelo internauta. Em alguns sistemas, o pesquisador indica no local de "opções de busca" que deseja que a expressão seja considerada como uma frase. Uma vez indicado este tipo de pesquisa, não é necessário colocar a frase entre aspas.

## • O que é pesquisa por campo?

Considerando-se que uma página *web* é feita de muitas partes, tais como o título; URL, texto da página; *links* da página; imagens na página, etc, e que estes

códigos HTML (ou suas variações), são "indexáveis", alguns sistemas de busca (tais como o Altavista e o Yahoo!) "lêem" as metatags da página e podem buscar as palavras da pesquisa no código HTML correspondente.

Exemplo

Uma página que tem o seguinte código:

```
<title>Power Searching</title>
```

Caso se faça uma pesquisa usando a sintaxe: **title: "Power searching"**, as páginas recuperadas devem conter, necessariamente no título (identificadas por código HTML), as palavras escritas acima (definidas como uma frase)

Outros exemplos de pesquisa pela metatag de título seriam:

title: marte

title: paisagens AND marte

title: "paisagens em marte"

- primeiro exemplo busca pela palavra "marte" em páginas de títulos
- segundo exemplo busca por ambas as palavras "marte" e "paisagens" em páginas de título
- terceiro exemplo busca pela frase exata "paisagens em marte" em páginas de título

O Altavista e o Yahoo! também recuperam pelo campo "autor", enquanto que o Altavista permite a busca por outras indicações de metatags, tais como:

- *applet* – busca arquivos executáveis em java
- *domain* – busca por domínio específicos
- *host* – busca por servidores de páginas web
- *image* – busca arquivos de imagens
- *link* – busca a identificação de determinado *link* de hipertexto definido, isto é, quantas páginas estão sendo remetidas (ou se remetem a uma específica)
- *text* – busca por um texto (incluído na meta tag `<text> </text>`, por exemplo)
- *title* – já exemplificada
- *url* – busca por determinado *Uniform Resource Locator* (ou, popularmente, por um endereço na web)

entre outras...

## • Elementos de meta-ferramentas de busca

As meta-ferramentas surgiram devido ao fato de que os sistemas disponíveis usam metodologias diferentes e a busca nem sempre alcança bons resultados. Para aqueles que não podem (ou não querem) consultar vários sistemas de busca, os chamados "metabuscaadores" podem ser uma solução. Estes sites reúnem, em uma única página, os resultados de pesquisas feitas em dezenas de serviços de busca simultaneamente. Com eles, o internauta insere sua estratégia de busca (palavras-chave e termos), e o sistema cuida do resto.

O metabuscador retransmite a mesma pesquisa para os sites cadastrados (Altavista, Yahoo!, Google, entre outros). Ao reunir as respostas que cada um destes envia, eliminam-se os resultados repetidos. Há metabuscas para todos os tipos de pesquisa, desde os famosos arquivos MP3, até para *softwares*, receitas culinárias, números de telefone, entre outras informações. Contudo, nenhum destes sites faz milagres, dependendo do assunto e do tipo de resultado esperado, os sistemas de busca comuns ainda são mais eficientes. Isto acontece porque os comandos especiais para refinar uma pesquisa (os operadores booleanos e a "matemática" de busca) variam de sistema para sistema.

Assim, muitos metabuscas não conseguem obter resultados focados – especialmente quando se utilizam palavras compostas ou frases (que acabam produzindo repostas extensas e imprecisas)

Exemplos de metabuscadores:

Miner

<http://www.miner.bol.com.br>

Oferece 38 sistemas de busca em sete categorias: compras, lazer, notícias, informática, banco de dados, música e pessoas. Pesquisa em quatro *sites* de busca no Brasil e seis no exterior, simultaneamente.

Bluebook

<http://www.bluebook.com.br>

Consulta simultaneamente 16 sites de busca no Brasil, 16 nos EUA, 16 em Portugal, 13 na Europa e 13 no resto do mundo, além de sistemas de busca especializados. Seu principal problema é não reunir os resultados em uma única tela, abrindo janelas diferentes para cada resultado.

Buscossauro Matrix

<http://www.buscossauro.com.br>

Pertence ao provedor Matrix. Reúne resultados de 11 sistemas. Funciona melhor com o Internet Explorer.

Jarbas

<http://www.jarbas.com.br>

Oferece pesquisas focadas no Brasil ou em todo o mundo. Apresenta opção para MP3 e um serviço que armazena os resultados da busca que mais interessaram ao internauta, criando um *bookmarks* (favoritos) *online*, que pode ser consultado de qualquer micro.

Metacrawler

<http://www.metacrawler.com>

Site em inglês que busca páginas da Internet, MP3, imagens, mensagens de grupos de discussão e leilões. Utiliza 13 sistemas de busca simultaneamente. Permite a personalização das páginas e da metodologia de busca, focando regiões específicas.

## • Bibliografia

ACKERMANN, Ernest. *Searching and researching on the Internet and the World Wide Web*: for students, teachers, librarians, and folks who want to stay in the know. Disponível em: <http://www.webliminal.com/search/index.html>. 1999. Acesso em: 18.05.00

CONTE JR, Ron. Faróis digitais. *Internet World*, v. 1, n. 9, maio 1996, p. 51-56

BUENO, André. Busca rápida para quem não quer mergulhar em águas profundas. *Jornal da Tarde*, 29 de junho de 2000. (Informática, p. 1D)

DENEGA, Marcos Antonio. *Como pesquisar na Internet: vá em frente e aventure-se já!* São Paulo: Berkeley, 2000. 217 p.

SAIBA usar a ferramenta certa para busca rápida. *Jornal da Tarde*, 29 de junho de (Informática, p. 10D)

SEARCH Engine Watch. Disponível em: <http://www.searchenginewatch.com> Acesso em: 18.06.99

SITE criado por estudantes é a nova sensação. *Jornal da Tarde*, 13 de julho de 2000 (Informática, p. 3D)

VAUGHAN-NICHOLS, Steven. Pesquisa a jato. *Internet World*, v. 2, n. 23, p. 88-91

VENDITTO, Gus. Quem procura, acha. *Internet World*, v. 1, n. 9, maio 1996, p. 71-82