



A pesquisa de informação na World Wide Web

Nelson Gonçalves
nelson@esev.ipv.pt

Resumo

Actualmente, surgem várias propostas de estratégias e actividades baseadas na World Wide Web (WWW) para implementar em contexto educativo. Grande parte destas propostas assenta sobre as competências de pesquisa na WWW dos alunos e do professor. Deste modo, emerge a necessidade de se conhecer de forma mais aprofundada o modo de funcionamento dos diferentes serviços de pesquisa actualmente disponibilizados.

[1. A pesquisa na WWW e a Educação](#)

[2. Os serviços de pesquisa na WWW](#)

[2.1 Directórios](#)

[2.2 Motores de Busca](#)

[2.3 Metabusca e Motores de Busca Híbridos](#)

[3. Formas básicas de pesquisar com termos ou palavras-chave](#)

[4. Conclusões](#)

[Referências](#)

[Apêndices 1](#)

1. A pesquisa de informação na WWW e a Educação




Com a World Wide Web, o espaço de aprendizagem alarga-se e reforça-se, estando a partilha do conhecimento cada vez menos limitada ao espaço da escola: “The Web permits the sharing of information as personal as one person to another or as public as one person to the entire world,”(Crossman, 1997:23).

Através da WWW, o aluno pode agora aceder a imensa informação e confrontar-se com diferentes perspectivas, actuais e clássicas, sobre os mesmos fenómenos. A utilização sistemática desta ferramenta permite implementar uma abordagem holística, multi e interdisciplinar, onde a aprendizagem resulta sobretudo de uma contribuição activa do aluno na construção do seu próprio conhecimento.

Existem várias formas de integrar a WWW, e num sentido mais a lato a Internet, no processo de ensino-aprendizagem de forma a tirar a máxima vantagem do seu potencial educativo. O trabalho colaborativo, a resolução de problemas e o trabalho de projecto são três exemplos de estratégias que com a WWW assumem um papel destacado e cada vez mais renovado. Independentemente da estratégia escolhida ou da actividade que o professor pretende implementar, a pesquisa de informação na WWW representa uma competência essencial tanto para alunos como para o próprio professor.

As *Webquests* são um exemplo, de entre vários, de como a integração da WWW na aprendizagem necessita da criação e desenvolvimento de novas competências na área da pesquisa. Para Dodge (1997) “A WebQuest is an inquiry-oriented activity in which some or all of the information that learners interact with comes from resources on the internet (...).” Consequentemente, a capacidade de pesquisa e localização destes recursos é fundamental.

Todavia, a WWW não é apenas uma magnífica fonte de documentos e informação, qual representação virtual da biblioteca de Córdova do califa El-Hakam II ou da Encyclopédie de Diderot e D’Alembert, mas antes uma ferramenta de renovação da educação onde a informação não constitui um fim em si mesmo mas sim um meio de reestruturação, de aprendizagem, onde o sujeito não se pretende reprodutivo mas produtivo, dinâmico, construtivo e motivado.

Urge, assim que no contexto da utilização educacional da pesquisa na WWW se tenha em consideração o aviso de Maddux (1997): “Although search engines represent a step forward in improving the teaching and learning potential of the World Wide Web, they are less than perfect”(p. 295). Conhecer o modo de funcionamento, potencialidades e limitações dos diferentes serviços de pesquisa na WWW é condição essencial para que seja possível integrar a pesquisa na WWW na aprendizagem em todos os contextos e ao longo da vida de forma cada vez mais eficaz. 




2. Os serviços de pesquisa na WWW

Actualmente, existem disponibilizados na WWW bastantes serviços de pesquisa. Todos eles visam um objectivo comum: permitir ao seu utilizador localizar informação. Contudo, se os analisarmos de forma mais cuidada, cedo constatamos que existem diferenças importantes entre esses mesmo serviços.

São sobretudo duas as características em que é possível assentar a diferenciação: o âmbito no qual incide a pesquisa disponibilizada (pode ser um serviço especializado ou generalista e pode ter um âmbito internacional, nacional ou regional) e o modo de funcionamento do próprio serviço.

Se sobre a primeira característica não há, de momento, muito a afirmar, dado que depende dos objectivos das pessoas ou instituição que disponibiliza o serviço, o mesmo não acontece com as características relacionadas com o seu modo de funcionamento.

Antes de diferenciarmos os vários tipos de funcionamento, talvez seja útil descrever o processo comum a todos eles. Ao contrário do que muitos utilizadores provavelmente consideram, não existe nenhum serviço de pesquisa que, com os termos e comandos introduzidos, circule por todas as páginas da WWW procurando a informação desejada. Na realidade, o serviço apenas pesquisa na sua base de dados e esta é construída de forma totalmente independente da pesquisa do utilizador. Os resultados apresentados ao utilizador não são fruto de nenhuma *viagem* específica na WWW mas apenas os encontrados nessa mesma base de dados.

A distinção essencial entre os dois grandes grupos em que se dividem os serviços de pesquisa da WWW, Directórios e Motores de Busca, incide justamente na natureza dessas bases de dados e na forma como as mesmas são construídas. 

2.1 Directórios

No início, os Directórios eram criados por indivíduos que localizavam, reuniam e classificavam em categorias endereços de documentos (URL – Uniform Resource Locator). Actualmente, há *software* capaz de realizar estas mesmas tarefas de forma mais rápida e eficiente. Todavia, continua a existir um elemento humano preponderante na construção das descrições dos documentos e na sua avaliação, recomendação e categorização.


Geralmente os registos que constituem a base de dados onde o Directório pesquisa, por vezes também designado por catálogo, são compostos por título, endereço (URL), breve descrição do seu conteúdo e uma série de palavras-chave relacionadas com o mesmo.



Há ainda a referir o facto de a quase totalidade dos Directórios permitir ao utilizador o envio de informações ou criação de registos relativos a documentos da WWW, sendo tal posteriormente verificado e validado. Esta característica realça mais uma vez a importância do factor humano na construção do serviço de pesquisa de tipo Directório.

A principal vantagem associada aos Directórios é o facto de toda a informação pesquisável estar categorizada. Este modo de organização confere-lhe um elevado grau de *amigabilidade* e permite a realização de pesquisas hierárquicas ou em árvore, limitando-se o utilizador a seleccionar as directorias (temas gerais) e as subdirectorias (temas mais específicos) que correspondem às suas preferências e objectivos.

Por outro lado, as principais desvantagens estão relacionadas com a quantidade diminuta de informação pesquisável, relativamente à totalidade de documentos existentes na WWW, e a rápida desactualização dos seus registos, pois um documento pode facilmente mudar de endereço (URL), ou até de conteúdo, sem que o seu registo seja actualizado.

Webster e Paul (1996) sintetizam as desvantagens dos Directórios da seguinte forma: "Because subject trees rely on humans for their overall design and maintenance, they typically provide links to a smaller number of documents than most automated search engines. The results of browsing, or even searching an index to a subject tree, are likely to be incomplete. For these reasons, subject trees do not always do the job and it is usually advisable to supplement a subject tree search with a visit to one or more search engines". 

2.2 Motores de Busca

O termo Motor de Busca é muitas vezes utilizado para designar realidades diferentes. No presente texto surge com o significado de motor de pesquisa e enquanto tipo de serviço cujo funcionamento é totalmente suportado por software próprio.

Todos os Motores de Busca disponibilizados na WWW são constituídos por três elementos essenciais. Em primeiro lugar, o *robot*, por vezes também designado por *crawler* ou *spider*, um programa que automaticamente percorre toda a estrutura hipertexto da WWW, identificando todos os elementos de texto contidos nos documentos que encontra e continuando o seu percurso, utilizando as hiperligações existentes nesses mesmos documentos. Geralmente, o *robot* retorna a todos os documentos já visitados com uma periodicidade mensal ou bimensal, tentando detectar alterações.



Num segundo momento, surge o *index*, por vezes também designado por catálogo. Este é um gigantesco repositório constituído por todos os elementos que o *robot* encontrou desde o URL ao próprio texto contido no documento. Nesta base de dados, existe como que uma cópia virtual de todas as páginas visitadas pelo *robot*.

Por fim e em terceiro lugar, surge o Motor de Busca disponibilizado ao utilizador. Este vai efectuar pesquisas na base de dados, tentando encontrar a informação desejada pelo seu utilizador, e apresentar os resultados obtidos de forma hierarquizada, de acordo com os critérios de relevância para o qual foi programado. Os dois métodos mais comuns de hierarquizar os resultados passam por dar relevância aos documentos de acordo com o número de vezes que os termos de pesquisa aparecem ou com a proximidade dos mesmos em relação ao topo dos documentos.

Atente-se no facto de o Motor de Busca considerar nas suas pesquisas todos os elementos contidos nos documentos e não apenas a sua descrição ou palavras-chave, como na generalidade dos Directórios. [^](#)

2.3 Metabusca e Motores de Busca Híbridos [^](#)

No que respeita ao modo de funcionamento dos serviços de pesquisa da WWW, resta-nos ainda referir três variantes: a Metabusca, os Motores de Busca Híbridos e os Portais.

Em relação à Metabusca, esta não constitui verdadeiramente um modo de funcionamento de um serviço de pesquisa. Trata-se de um serviço que permite activar pesquisas simultaneamente em Motores de Busca e Directórios diferentes. Estes serviços lançam os mesmos termos, palavras-chave, e comandos de pesquisa em diferentes Motores de Busca e Directórios, apresentando, de uma só vez, todos os resultados, hierarquizados consoante os seus próprios critérios de relevância e identificando os devidos créditos da pesquisa.

Os Motores de Busca Híbridos são, na maior parte das vezes, Motores de Busca aos quais se associou um serviço de Directório. Procura-se desta forma aliar a potencialidade de pesquisa dos Motores de Busca e a organização dos Directórios. O *index* é construído pelas viagens dos *robots* e o Directório pelas descrições enviadas pelos utilizadores ou pela própria equipa que disponibiliza o serviço.

O modo de funcionamento dos Motores de Busca é, muitas das vezes, o único aspecto que todos têm em comum, como referem Webster e Paul (1996): "Search engines vary according to the size of



the index, the frequency of updating the index, the search options, the speed of returning a result set, the result set presentation, the relevancy of the items included in a result set, and the overall ease of use”.

A opção do utilizador tem pois de considerar os vários aspectos, sendo que um dos mais importantes diz respeito às especificidades de cada um destes serviços. Tal como sugere a Kansas City Public Library (1999): “Subject guides are fine for browsing general topics, but for specific information use a search engine”.

Por outro lado, os objectivos da pesquisa do utilizador e a sua própria competência na área em que pretende pesquisar são também aspectos essenciais que condicionam os resultados obtidos: a clareza e exactidão dos termos de pesquisa facilitam a exactidão dos resultados. ▲

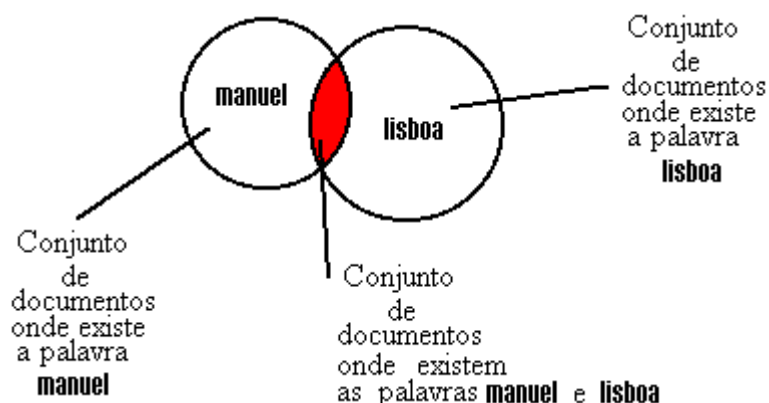
3. Formas básicas de pesquisar com termos ou palavras-chave

Existem três formas básicas de pesquisar com termos, com palavras-chave, em Motores de Busca ou Directórios: pesquisa com todas as palavras (Match All), pesquisa com qualquer das palavra (Match Any) e pesquisa com frase exacta (Exact Phrase).

No primeiro caso, se digitarmos dois termos, o serviço de pesquisa irá apresentar nos seus resultados os documentos onde encontrou esses mesmos dois termos incluídos. Não tem de existir qualquer relação ou ordem entre os termos. Assim, exemplificando, se levarmos a cabo uma pesquisa deste género com os termos *manuel* e *lisboa*, um dos resultados possíveis seria um documento com um rol de nomes onde existisse alguém chamado António *Lisboa* e *Manuel* Silva.

Gráfico 1

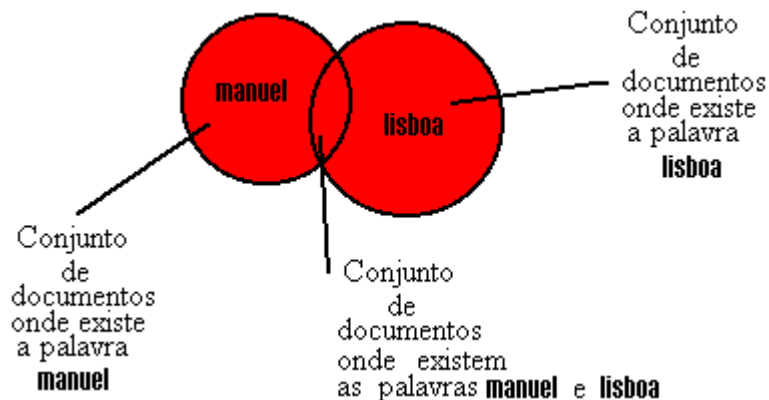
Match All - Pesquisa com todas as palavras



Na pesquisa de tipo Match Any, com os mesmo dois termos, o serviço de pesquisa irá apresentar nos resultados todos documentos onde exista pelo menos um deles. Geralmente, os resultados estão hierarquizados de modo a que aqueles documentos que possuam simultaneamente os dois termos sejam apresentados no início da lista. Para utilizarmos o exemplo anterior, podemos afirmar que nos nossos resultados irão surgir todos os documentos que contenham a palavra *manuel* ou *Lisboa*.

Gráfico 2

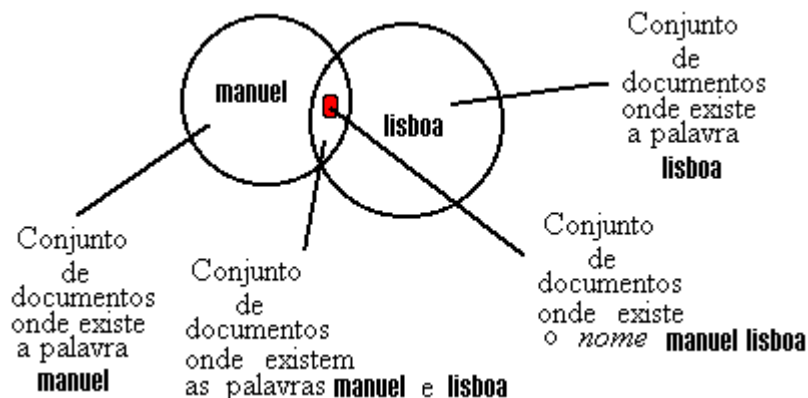
Match Any - Pesquisa com qualquer uma das palavras



A última forma básica de pesquisar, pesquisa com frase exacta (Exact Phrase), caracteriza-se por apresentar ao utilizador apenas os documentos onde os termos existam na ordem exacta com que foram digitados. Se digitarmos *manuel lisboa*, o resultado da pesquisa será o conjunto de documentos onde exista essa sequência exacta de palavras. Ficam assim excluídos documentos com a sequência *manel lisboa* (*manuel* é um termo diferente de *manel*) ou *manuel foi a lisboa*.

Gráfico 3

Exact Phrase - Pesquisa com frase exacta



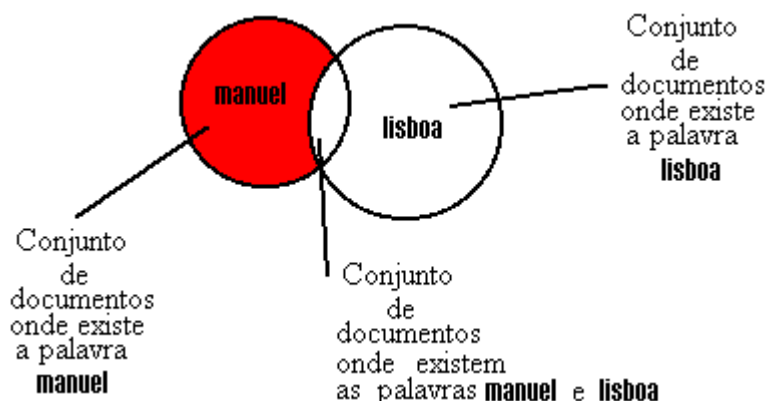
Em muitos serviços de pesquisa, a opção do utilizador por uma das formas acima descritas é feita através de um menu próximo da caixa de texto onde se pode digitar os termos de pesquisa. Quando este menu não existe, a opção é tomada através da introdução de operadores, juntamente com os termos, na já referida caixa de texto.

Através da introdução dos operadores também é possível excluir dos resultados documentos que contenham determinadas palavras. Esta possibilidade reforça a exactidão dos resultados. É possível

indicar ao serviço de pesquisa que pretendemos todos os documentos com os termos *manuel* e *lisboa* e deste conjunto excluir os documentos que possuam no seu conteúdo o termo *maria*, ou todos os documentos com o termo *manuel* e excluir destes resultados aqueles que possuam o termo *maria*.

Gráfico 4

Pesquisa com exclusão de uma palavra



Podemos ainda mencionar o operador NEAR ou a utilização de () para construir pesquisas mais complexas. O NEAR, que apenas pode ser utilizado nos serviços que suportam pesquisa por booleanas, permite indicar ao serviço de pesquisa que dois termos devem estar perto um do outro (exemplo: *manuel* NEAR *lisboa*). Geralmente, nos serviços que permitem este operador, o grau de proximidade já está predefinido. Não obstante, alguns destes serviços permitem ainda definir o grau de proximidade. Se quisermos indicar que os termos, no máximo, só podem distar entre si 25 palavras, deveríamos digitar *manuel* NEAR/25 *lisboa*.

Os parêntesis, na maior parte dos casos, também só podem ser utilizados nos serviços de pesquisa que permitam booleanas. Nesse caso, deveríamos digitar *lisboa* AND (*manuel* OR *maria*) para obter os documentos que contivessem o termo *lisboa* e pelo menos um dos outros dois.


Poderíamos ainda mencionar a pesquisa no título dos documentos (texto contido entre os comandos <title> e </title> do código HTML do documento), no próprio endereço (URL) do documento, a pesquisa de documentos que possuam hiperligações apontadas para um outro endereço (URL) ou a utilização de asterisco (* - *wildcard*). Contudo, parece-nos que essas possibilidades já pertencem a um nível mais avançado de pesquisa que a esmagadora maioria dos utilizadores não necessita de possuir para conseguir encontrar a informação desejada. ⚡



4. Conclusões

O número actual de documentos da WWW é extraordinariamente elevado e o seu ritmo de crescimento não parece abrandar. Apesar da melhoria relativa que se tem verificado nos serviços de pesquisa para a WWW, estes abrangem uma percentagem cada vez menor do total de documentos existentes. Paralelamente, a sua precisão parece diminuir, sendo comum o utilizador obter numa pesquisa milhares de documentos inúteis. O utilizador, apesar das melhorias nos serviços de pesquisa, tem grandes dificuldades em conseguir localizar a informação de que necessita. Esta dificuldade é agravada, na maior parte dos casos, pela utilização incorrecta ou desadequada dos serviços de pesquisa existentes e pelo desconhecimento do seu modo de funcionamento.

Actualmente, surgem várias propostas de estratégias e actividades baseadas na World Wide Web para implementar em contexto educativo. Grande parte destas propostas assenta sobre as competências de pesquisa na WWW dos alunos e do professor. Deste modo, emerge a necessidade de se conhecer de forma mais aprofundada o modo de funcionamento dos diferentes serviços de pesquisa actualmente disponibilizados.

Podemos distinguir dois grandes tipos de pesquisa: a hierárquica ou em árvore e a pesquisa com termos ou palavras-chave. Existem ainda dois grande grupos de serviços para essas pesquisas (Directórios e Motores de Busca). Estes dois grupos distinguem-se entre si pelo seu modo de funcionamento, sendo que ambos podem aumentar a sua eficácia através da utilização de operadores. Os operadores são comandos digitados que permitem implementar três estratégias de pesquisa diferentes (Match All, Match Any e Exact Phrase), assim como aumentar a eficácia dos resultados através de várias funcionalidades (exclusão de termos, pesquisa complexa/composta, proximidade de termos, etc). 



Referências

Barlow, Linda (1999). *The Spider's Apprentice*.

Disponível em <http://www.monash.com/spidap.html> [20/2/2000].

California University Library (1999). *Finding Information on the Internet: A Tutorial*.

Disponível em <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html> [24/3/2000].

Chakrabarti, Soumen; Dom, Byron; Gibson, David ; Kleinberg, Jon M.; Kumar, S. Ravi; Raghavan, Prabhakar; Rajagopalan, Sridhar; Tomkins, Andrew (1999). "Hypersearching the Web". *Scientific American*, Junho de 1999.

Disponível <http://www.sciam.com/1999/0699issue/0699raghavan.html> [5/1/2000].

Crossman, David M. (1997). "The Evolution of the World Wide Web as an Emerging Instructional Technology Tool". In Khan, Badrul H. (Ed), *Web-Based Instruction*. New Jersey: Educational Technology Publications, pp. 19-23.

Cyberatlas (2000).

Disponível em <http://cyberatlas.internet.com> [28/1/2000].

Dodge, Bernie (1997). *Some Thoughts About Webquests*.

Disponível em http://edweb.sdsu.edu/courses/edtec596/about_webquests.html [12/3/2000]

DomainStats (2000).

Disponível em <http://www.domainstats.com> [28/1/2000].

Fondren Library – Rice University (2000). *Internet Searching Strategies*.

Disponível em <http://riceinfo.rice.edu/Fondren/Netguides/strategies.html> [10/3/2000].

Forsyth, Ian (1998). *Teaching and Learning Materials and the Internet*. London: Kogan Page.

Global Reach (2000). *Global Internet Statistics*.

Disponível em <http://www.euromktg.com/globstats/> [28/1/2000].

Graphics, Visualization, & Usability Center (1998). *GVU's 10th WWW User Survey*.

Disponível em http://www.gvu.gatech.edu/user_surveys/survey-1998-10/ [20/1/2000].



Habib, David P. e Balliot, Robert L. (1999). *How to Search the World Wide Web: A Tutorial for Beginners and Non-Experts*.

Disponível em <http://204.17.98.73/midlib/tutor.htm> [1/3/2000].

Harris, Charlie (2000). *Internet Search FAQ*.

Disponível em <http://www.purefiction.com/pages/resfaq.htm> [15/3/2000].

Kansas City Public Library (1999). *Introduction to Search Engines*.

Disponível em <http://www.kcpl.lib.mo.us/search/srchengines.htm> [12/3/2000].

Koster, Martijn (s/d). *The Web Robots FAQ*.

Disponível em <http://info.webcrawler.com/mak/projects/robots/faq.html> [13/1/2000].

Lawrence, Steve e Giles, C. Lee (1999). "Accessibility of information on the web". *Nature*, vol. 400, pp. 107-109, Julho de 1999.

[exemplar enviado pelos autores por correio electrónico]

Maddux, Cleborne D. (1997). "Web-Based Search Engines". In Khan, Badrul H. (Ed), *Web-Based Instruction*. New Jersey: Educational Technology Publications, pp. 295- 298.

McKim, Geoffrey W. (1996). *Internet Research Companion – Making the right connections and discoveries*. Indianapolis: Que Education & Training.

Search Engine Watch (2000).

Disponível em <http://searchenginewatch.com/> [2/2/2000].

Tilburg University Library (2000). *Searching the world wide web: a basic tutorial*.

Disponível em <http://cwis.kub.nl/~dbi/english/instruct/www/indexuk.htm> [4/3/2000].

Twidale, Michael B.; Nichols David M.; Smith, Gareth; Trevor, Jonathan (1995). *Supporting Collaborative Learning during Information Searching*.

Disponível em <http://www.comp.lancs.ac.uk/computing/research/cseq/projects/ariadne/docs/cscl95.html>

[17/3/2000].

University of Bristol Computing Service (1999). *Searching for information on the World Wide Web*.

Disponível em <http://www.cse.bris.ac.uk/help/docs/web-r3/web-r3.htm> [19/3/2000].

Webster, Kathleen e Paul, Kathryn (1996). *Beyond Surfing: Tools and Techniques for Searching the Web*.


 Disponível em <http://magi.com/~mmelick/it96jan.htm> [15/3/2000].

Apêndice 1

Quadro 1

	Operadores	
	<i>Símbolos</i>	<i>Booleanas</i>
Todas as palavras <i>Match All</i>	+ +manuel +lisboa	AND manuel AND lisboa
Qualquer das palavras <i>Match Any</i>		OR manuel OR lisboa
Frase Exacta <i>Exact Phrase</i>	“ “ “manuel lisboa”	
Exclusão de Termos	- manuel –maria	NOT ou AND NOT manuel NOT maria manuel AND NOT maria
Combinar operadores	“manuel lisboa” –maria +manuel +lisboa –maria	manuel AND lisboa AND NOT maria
Outros operadores * () NEAR		navega* manuel AND (lisboa OR maria) (manuel AND lisboa) OR (maria AND lisboa) manuel NEAR lisboa manuel NEAR/10 lisboa