

Mecanismos de Busca Acadêmica: uma Análise Quantitativa

Diego Buchinger¹

Gustavo Andriolli de Siqueira Cavalcanti¹

Marcelo da Silva Hounsell¹

Resumo: Os mecanismos de busca consolidaram a pesquisa por informação na Internet, enquanto que os mecanismos de busca acadêmicos (MBAs) são hoje os principais meios para alcançar fontes científicas como artigos de eventos ou periódicos. Ao invés de gastar horas em uma biblioteca, os MBAs ganham em popularidade pela facilidade de automatizar uma busca e acessar inúmeras bases de dados simultaneamente. Contudo, o caráter incremental das bases de materiais científicos e dos diversos MBAs existentes impõem um desafio aos pesquisadores, que é determinar um MBA que melhor atenda aos seus anseios de pesquisa. Nesse contexto, uma análise quantitativa dos recursos providos pelos MBAs foi realizada com objetivo de determinar quais MBAs são os melhores para facilitar as buscas por conteúdo científico, considerando recursos de busca, de refinamento e auxiliares. Dentre um grupo de 40 MBAs iniciais, sete se apresentaram $\sigma = 0,5$ acima da média nos critérios utilizados, na seguinte ordem: 1º) Web of Knowledge; 2º) Engineering Village; 3º) Scopus SciVerse; 4º) IEEE Xplore; 5º) ACM Digital Library; 6º) Science Direct; e 7º) Springer Link. Estes MBAs deveriam ser melhor explorados pelos pesquisadores pois tem inúmeros recursos capazes de facilitar sobremaneira e elevar a qualidade do trabalho de pesquisa.

Palavras-chave: Mecanismos de Busca Acadêmica. Metodologia da Pesquisa. Pesquisa Científica.

Abstract: *Search engines have consolidated the search for information on the Internet, whereas academic search engines (ASEs) are the primary means to obtain scientific papers from journals and proceedings. Rather than spending many hours in a library, ASEs gain in popularity because they make it easier to automate a search and access various databases simultaneously. However, the variety of existing ASEs and the incremental character of the scientific databases pose a challenge to researchers in determining which ASE best meets their research desires. In this context, a quantitative analysis on ASE's features was performed in order to determine the ASEs that have the best conditions to facilitate the research for scientific content, considering search, refinement and auxiliaries' features. Among a group of 40 ASEs, seven had presented $\sigma = 0.5$ above the used criteria average, in the following order: 1st) Web of Knowledge; 2nd) Engineering Village; 3rd) Scopus SciVerse; 4th) IEEE Xplore; 5th) ACM Digital Library; 6th) Science Direct; and 7th) Springer Link. These ASEs should be better explored by the researchers because they provide a wide variety of resources that contribute to facilitate and improve the quality of the research work.*

Keywords: *Academic Search Engines. Research Methodology. Scientific Research.*

1 Importância dos mecanismos de busca acadêmica (MBAs)

Até por volta dos anos 1998 as pesquisas eram conduzidas de forma manual em enormes bibliotecas. Atualmente se tem acesso a uma quantidade enorme de materiais sem precisar ir ao templo do conhecimento, basta um acesso à Internet. Vive-se portanto a “era da informação” na qual deve-se gerar conhecimento. A busca por conhecimento pode ser entendida como ciência, e ela é incremental: é necessária uma base sólida de fundamentação para se realizar uma investigação e alcançar resultados. Neste mesmo sentido, existe uma frase que se costuma

¹Programa de Pós-Graduação em Computação Aplicada (PPGCA), Universidade do Estado de Santa Catarina (UDESC) - Joinville - Brasil
Email: {diego.buchinger@outlook.com, gasc@joinville.udesc.br, marcelo.hounsell@udesc.br}

atribuir a Isaac Newton, que diz: “Se vi mais longe do que outros é porque estava me apoiando em ombros de gigantes”. Isso reforça a necessidade da busca por conhecimentos e experiências como referências. Com a utilização da Internet, a ciência foi beneficiada e está em constante evolução, facilitando as trocas de experiências, a apresentação de teorias, testes, resultados e, enfim, a publicação de conhecimentos em forma de artigos. A Internet mudou completamente a maneira em que se reúnem informações, requerendo apenas alguns toques no teclado [1].

Para um pesquisador é importante encontrar formas eficientes e, principalmente disponíveis, de acesso às informações relevantes a sua pesquisa. Todavia, com o crescimento do número de dados disponíveis na Internet, incluindo o número de publicações científicas, as buscas na Internet começaram a se tornar mais complexas. Nesse cenário, vários mecanismos de busca genéricos (que mantêm o foco no perfil da maioria dos usuários da Internet, cobrindo todo tipo de informação) e específicos (voltados para uma área específica, como por exemplo, a academia) foram desenvolvidos para facilitar o acesso aos dados [2]. Voltado para o meio científico existem os mecanismos de busca acadêmica (MBAs), que facilitam o acesso às publicações científicas, com a possibilidade do uso de diversos filtros de ajuda para buscas mais concisas [3]. Em seu estudo, Rieger [4] justificou a necessidade atual do uso de MBAs para pesquisas acadêmicas, evidenciando a nova realidade nas universidades.

Para o contexto deste artigo, entende-se que “motor de busca” é um algoritmo eficiente para vasculhar bases de dados e que “mecanismo de busca” é a composição de site, motor de busca, recursos de interface e conjunto de bases de dados que cobrem uma determinada área do conhecimento, visando facilitar a identificação de materiais específicos e relevantes. Nesse contexto, surge a pergunta: quais são os MBAs mais adequados para facilitar o acesso a materiais científicos? Para responder esta pergunta, algumas análises são possíveis, como, por exemplo, pela preferência de um grupo de usuários ou pela aparência e usabilidade dos MBAs. Entretanto, este estudo adotou critérios quantitativos para análise, focado nas áreas das Ciências Exatas, da Terra e Engenharias, por ser uma população que valoriza informações objetivas. Assim, este estudo foi proposto com o objetivo de elencar os MBAs que apresentam as melhores condições para facilitar as buscas por conteúdo científico, com foco em artigos. Trabalhos relacionados foram estudados para auxiliar no processo de verificação, identificando características que foram consideradas nesta análise. Nesse sentido, realizou-se um estudo comparativo de 40 MBAs (veja a lista no Apêndice) e suas respectivas funcionalidades.

A estrutura deste artigo foi organizada da seguinte forma: a Seção 2 apresenta uma revisão bibliográfica. A Seção 3 discute a filtragem inicial dos MBAs escolhidos, detalhando os critérios utilizados e descrevendo considerações advindas da experiência da busca de dados. A Seção 4 apresenta a segunda fase de classificação dos MBAs e discute os critérios utilizados. A Seção 5 discute o resultado final da avaliação. A Seção 6 conclui o artigo.

2 A busca pelo melhor mecanismo de busca

Diversos autores já realizaram algum tipo de mensuração ou avaliação de MBAs. Foi constatado que os MBAs mais avaliados são: Web of Knowledge, Scopus e Google Scholar. Os estudos apresentados a seguir, além de servirem como uma revisão bibliográfica, também foram úteis na composição dos critérios de avaliação adotados nesta pesquisa.

Brophy and Bawden [5] compararam o mecanismo de busca (convencional) do Google com as bibliotecas universitárias. A metodologia utilizada foi quantitativa nos aspectos de revisão de documentos encontrados, precisão e sobreposição dos resultados, e qualitativa durante a verificação dos resultados encontrados. Descobriu-se que o mecanismo Google é melhor em quesitos de cobertura e acessibilidade do que os sistemas de bibliotecas, porém é pior na qualidade do material.

Segundo Meier e Conkling [6], a ferramenta de pesquisa Google Scholar, uma variação da ferramenta Google específica para pesquisas acadêmicas, é apropriada para pesquisas bibliográficas. Eles avaliaram a eficiência do Engineering Village, ferramenta tradicional para pesquisa na área de engenharia, e do Google Scholar, comparando a cobertura das pesquisas entre elas (i.e., o quanto os resultados de uma também aparece nos resultados da outra). Através de pesquisas em diferentes áreas de engenharia, divididas por décadas entre 1950 até 2000, foi observado que a cobertura e crescimento de ambos os MBAs era progressivo, sendo que a ferramenta Google Scholar apresentou um crescimento na cobertura de 33% (em média) na década de 1950, e 88% na década de 2000, em relação ao MBA Engineering Village.

Meho e Yang [7] examinaram o uso dos mecanismos Web of Knowledge, Scopus e Google Scholar buscando, entre outros aspectos, descobrir similaridades e diferenças em algumas de suas características como: período de cobertura, tipos de documentos, idiomas, cobertura de temáticas, e identificação de pontos fortes e fracos. Os resultados mostraram que Scopus e Web of Knowledge, embora tenham dados em comum, são complementares – há um percentual significativo de documentos que são encontrados em um mecanismo, mas não no outro – e que o Google Scholar retornou a maior quantidade de documentos únicos (i.e., que não foram encontrados nos outros MBAs), porém menos significativos e de menor qualidade do que as duas outras ferramentas.

Outra avaliação dos mecanismos Web of Knowledge, Scopus e Google Scholar mostrou que existe uma sobreposição considerável entre o primeiro e o segundo, enquanto que o terceiro apresentou uma cobertura muito maior [8]. Além disso, novamente os resultados indicaram que cada mecanismo possui os seus registros únicos que eram cobertos por nenhum dos outros mecanismos. Diversos fatores foram considerados na pesquisa, e em partes Jacso [8] criticou o Google Scholar por não permitir uma busca avançada com especificações de determinados elementos chaves de publicações, por apresentar problemas com operadores booleanos, por apresentar resultados inconsistentes (e.g., resultados em uma pesquisa considerando os últimos dez anos retornou mais registros do que uma pesquisa considerando os últimos vinte anos) e redundantes.

A fim de comparar as vantagens e desvantagens nos mecanismos de busca PubMed, Scopus, Web of Knowledge e Google Scholar, Falagas et. al. [9] realizaram uma pesquisa avaliando a cobertura e utilidade destas ferramentas para a recuperação de informação biomédica, e a análise de citações. Como resultado destacou-se que o Scopus foi o mecanismo que indexava o maior número de *journals*, enquanto que o Scholar oferecia um maior potencial internacional de pesquisa (pois não se restringe apenas à materiais em inglês), e o Web of Knowledge continha os registros mais antigos. Além disso, enfatizou-se que o PubMed é principalmente voltado para a área médica e de ciências biomédicas, enquanto que os outros mecanismos estudados cobrem outras diversas áreas da ciência. Outra característica enfatizada pelos autores é a indexação de materiais em versões anteriores à publicação efetiva feita pelo PubMed e também pelo Scholar, mas não realizada pelo Scopus e Web of Knowledge.

Atsaros, Spinellis e Louridas [2] compararam mecanismos de busca genéricos e mecanismos de busca específicos para pesquisa acadêmica, avaliando-os com critérios de precisão – entre os 10 primeiros artigos, quantos são relevantes – e eficiência – relação entre o número de registros encontrados num mecanismo, e o total de registros encontrados em todos os mecanismos. Foram avaliados os mecanismos de busca genéricos Google e Yahoo!, e os mecanismos de busca específicos (da área de comércio eletrônico e da área acadêmica) ACM DL, Amazon, Barnes & Noble, IEEE Xplore, IMBD, Ingenta Connect, JSTOR, PubMed, Scopus e Springer. Verificou-se que Scopus, IEEE Xplore e PubMed foram os melhores mecanismos específicos, apresentando bons resultados de precisão e eficiência. Ficou notável também que os mecanismos genéricos, apesar de apresentarem numerosas indexações de conteúdos, não possuíam indexação para determinados conteúdos procurados. Entretanto, foi o mecanismo de busca Google que apresentou o melhor resultado geral. Vale salientar ainda que o ACM DL teve uma avaliação fraca porque nas pesquisas realizadas o mecanismo não encontrou 11 artigos do seu próprio banco de dados.

Mais recentemente, Turner [10] realizou uma pesquisa com alguns MBAs – IEEE Xplore, ACM DL, ScienceDirect, Web of Knowledge, Scopus, Google Scholar and CiteSeer^X – usados para buscas científicas pela comunidade de engenharia de software, e forneceu informações sobre eles com resultados superficiais e pouco abrangentes no sentido do foco desta pesquisa.

Utilizando uma abordagem diferente, Jansen e Spink [11] conduziram um estudo a fim de comparar os logs de cinco mecanismos de busca – Excite, Fireball, AltaVista, BWIE, e AtlltheWeb.com – para analisar como as pessoas estão fazendo as suas buscas. Assim como é importante identificar os melhores mecanismos de busca, também é importante conhecer os hábitos de busca dos usuários, a fim de descobrir se eles realmente utilizam tudo o que a ferramenta oferece, e caso não o façam, oferecer treinamento adequado para eles. Dentre as conclusões, enfatizou-se que: a interação entre pesquisador e mecanismo de busca se tornou menos complexa; o número de buscas e o número de termos nas buscas são equiparáveis com dados antigos; e o uso de operadores booleanos nas buscas foi muito variável e não pôde ser utilizado como modelo para prever um comportamento.

As pesquisas de Meho e Yang [7] reforçam a ideia de que cada MBA requer estratégias específicas para a obtenção de melhores resultados, assim como as pesquisas de Rieger [4] que descobriu através de observações que, embora as inúmeras ferramentas de pesquisas ofereçam várias funcionalidades que facilitam as buscas, os pesquisadores utilizam basicamente palavras-chave em suas pesquisas e raramente fazem uso de opções avançadas.

Tal fato pode justificar futuros trabalhos que abordem questões sobre ensino ou treinamento no uso dos MBAs.

Além de pesquisas diretamente relacionadas aos MBAs, existe também uma ferramenta gratuita específica para a avaliação destes mecanismos: o JISC ADAT² (JISC Academic Database Assessment Tool). Essa ferramenta de avaliação é disponibilizada em formato web e é focada principalmente para os bibliotecários, a fim de permitir a comparação de tamanho, cobertura e funcionalidades de alguns MBAs específicos [12]. Para exemplificar, através desta ferramenta é possível descobrir quantos títulos de publicações são oferecidos por um MBA específico, além de comparar dois MBAs com base em quantos títulos são únicos e quantos são sobrepostos entre eles. Esta ferramenta é útil, pois os MBAs são dependentes de várias bases diferentes; não existe um MBA que tenha o mesmo conjunto que outro, fortalecendo a hipótese de que certos MBAs são complementares. Zhao e Logan [13] por exemplo, verificaram que as bases Science Citation Index e ResearchIndex possuem apenas cerca de 10% de registros em comum.

Oppenheim, Morris e McKnight [14] perceberam que, apesar de haver muitos estudos envolvendo esses mecanismos, há pouca consistência no modo de avaliação deles. Considerando isso, sugerem métodos de avaliação alternativos que podem ser utilizados para padronizar comparações entre mecanismos de busca. Esses métodos poderiam enfatizar fatores humanos, escolha de consultas, sintaxe de consultas e os resultados. Além disso, na revisão bibliográfica destacam alguns dos critérios de avaliação mais utilizados pelos pesquisadores: cobertura, atualidade, relevância, sintaxe da busca, áreas temáticas, evolução dos mecanismos, tempo de resposta, disponibilidade de recursos do sistema, opções e filtros de busca, fatores humanos e de interface, e qualidade dos resumos.

As comparações entre mecanismos de busca mostradas consideraram apenas um conjunto pequeno de mecanismos e abordam uma análise qualitativa com critérios muito variados e não padronizados. Portanto, as comparações existentes, apesar de esclarecedoras em si, são muito restritas. Uma análise mais objetiva e abrangente envolvendo um número maior de mecanismo de busca se faz necessário. Acredita-se que quanto mais os pesquisadores conhecerem a variedade de mecanismos de busca e seus recursos, melhores serão suas buscas, obtendo informações melhores e mais atuais, refletindo no desenvolvimento das suas pesquisas. Este estudo tentou identificar, dentre um conjunto amplo de possibilidades, os MBAs que possuem os melhores recursos para facilitar pesquisas científicas nas áreas tecnológicas.

3 Seleção inicial

Existem vários MBAs disponíveis na web, sendo que muitos deles parecem ser pouco conhecidos e utilizados pela academia no Brasil. Foi considerada a inclusão de tais mecanismos para conhecê-los e, sobretudo, oferecer uma chance de comparação com os atuais mecanismos mais largamente conhecidos. Entretanto, é necessário executar uma filtragem inicial sobre eles a fim de reduzir o número de MBAs para uma análise posterior, mais apurada. Assim, foi escolhida uma lista de 40 MBAs³ e foram definidos critérios de inclusão e de classificação que atendem às características de tamanho, cobertura, disponibilidade e utilidade, com base em alguns dos critérios apresentados por Oppenheim, Morris e McKnight [14], conforme apresentado na Tabela 1. Vale ainda salientar que existe uma possível discussão sobre a classificação do Periódico da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Scholar Google como MBAs (ambos fazem parte da lista de MBAs escolhidos). Como ambos contêm muitas das funções características de um MBA, além de serem portais para outros MBAs, eles foram mantidos nesta pesquisa.

A fim de indicar quais MBAs possuíam as características de disponibilidade, cobertura e utilidade na Tabela 1, conforme descrito, foi utilizado o símbolo ● para aqueles que atendiam aos critérios definidos individualmente, e o símbolo ○ para aqueles que não os atendiam. Assim, utilizando os critérios de classificação e inclusão apresentados, os 40 MBAs iniciais foram classificados e filtrados entre os dias 3 a 8 de abril de 2013. Desta classificação seis MBAs (Aca. Pub. eJournal, ArnetMiner, GJS, Mendeley, NSDL e OCLC) foram descartados por não se tratarem efetivamente de MBAs e outro (SciDiver) foi descartado por não mais existir.

O primeiro critério de classificação utilizado, tamanho da base, foi mensurado com base em dois parâmetros: o número de materiais científicos (artigos, resumos, citações) divulgados (coluna Anunciado na Tabela 1) e,

²<http://adat.crl.edu/>

³Consulta feita em Fev/2013 no site Wikipedia – http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines

como alguns MBAs não informavam este dado, adotou-se a estratégia de mensurá-lo através do número de resultados obtidos por uma consulta genérica com termos básicos da língua inglesa (coluna Estimado), pois considera-se este idioma como a língua da ciência.

Tabela 1: Seleção inicial dos MBAs utilizando critérios de tamanho, cobertura, disponibilidade e utilidade.

MBA	Tamanho da Base (milhares)		Disponibilidade	Cobertura		Utilidade
	Anunciado	Estimado		Geo.	Acess.	
Scirus	545.000	564.237	●	●	●	●
Web of Knowledge	*	58.728	●	●	●	●
Scopus SciVerse	49.000	33.333	●	●	●	●
Periódicos da Capes	*	48.626	●	●	●	●
BASE	44.052	44.052	●	●	●	●
Socol@r	23.895	*	●	●	●	●
Engineering Village	2.791	14.377	●	●	●	●
Science Direct	11.642	11.352	●	●	●	●
Google Scholar	*	10.100	●	●	●	●
SciELO	9.697	279	●	●	●	●
Ingenta Connect	5.854	4.551	●	●	●	●
Springer Link	4.733	4.733	●	●	●	●
CiteSeer ^X	*	4.426	●	●	●	●
IEEE Xplore	3.437	*	●	●	●	●
CCSB	3.000	*	●	●	●	●
DBLP	2.246	*	●	●	●	●
ACM DL	*	2.106	●	●	●	●
inSpire HEP	*	998	●	●	●	●
WorldWideScience.org	*	204.528	●	●	●	○
Science.gov	200.000	10.443	●	●	●	○
Science Accelerator	*	37.844	●	●	●	○
Microsoft AS	*	31.195	●	●	●	○
NDL	*	30.476	●	●	○	*
Citeulike	6.767	*	●	●	●	○
ICI	4.992	*	○	*	*	*
Emerald	670	*	●	●	●	●
Pubget	*	399	●	●	●	●
SafetyLitv	375	379	●	●	●	●
AJO	155	*	●	●	●	●
Cogprints	4	*	●	●	●	○
arXiv	*	*	●	●	●	●
Airiti	*	*	*	*	○	*
SciFinder	*	*	*	*	○	*

* Não foi possível coletar este dado

O resultado da busca genérica foi definido como sendo o maior número de materiais científicos que o mecanismo de busca diz ter encontrado com base em pesquisas isoladas contendo os termos: “a”, “the”, “of”, “*” e “ ” (espaço em branco), sendo que o asterisco indica o caractere coringa do mecanismo de busca. Vale salientar que o operador booleano ‘OR’ não foi utilizado para agregar os resultados porque, como mencionado na revisão bibliográfica, alguns mecanismos possuem certos problemas com estes operadores. Para mostrar como a estratégia de utilização de uma busca genérica pode ser efetiva, pode-se observar a proximidade entre os valores anunciados e estimados que alguns dos MBAs apresentaram. Por outro lado, alguns mecanismos possuem algumas restrições que não permitiram as buscas genéricas realizadas. Por este motivo não foi possível estimar o tamanho de base de todos os MBAs considerados.

A questão da disponibilidade de acesso aos artigos foi mensurada com base no nível de acesso concedido

pela CAPES, visto que este órgão concede o acesso aos principais mecanismos pagos às universidades brasileiras.

O critério de cobertura foi avaliado como a cobertura geográfica (coluna Geo. da Tabela 1) das publicações contempladas nos mecanismos – i.e., se o mecanismo contempla documentos publicados em diferentes localidades – e a cobertura de acessibilidade (coluna Acess. da Tabela 1) através do idioma de apresentação do web site, que deve ser em inglês – a língua da ciência – ou em português – para o contexto nacional.

O critério de utilidade foi qualificado pelos tipos de materiais que o buscador disponibiliza. Considerou-se para este critério que, apesar de quanto mais materiais disponíveis melhor, é preciso que o mecanismo de busca não agregue materiais de baixa confiabilidade – tais como sites de notícias, blogs ou qualquer outro material que claramente não é avaliado por um corpo revisor – sendo considerados úteis apenas aqueles mecanismos que não buscam materiais desta natureza ou que permitam a sua filtragem. Numa busca científica, entende-se que o que é considerado útil são aqueles documentos que possuem uma validação apropriada, sendo revisados por um grupo de especialistas da área, tais como citações, artigos e periódicos.

Com a separação dos MBAs que apresentaram os critérios de inclusão, o que resultou numa lista de 22 MBAs, selecionou-se um novo critério de inclusão baseado no tamanho das bases. Este critério foi definido como sendo o tamanho de base – anunciado ou estimado – maior do que um milhão de documentos. Contudo, sendo que o MBA inSpire HEP apresentava um tamanho muito próximo ao critério de corte (998 mil documentos) este também foi incluído na lista dos MBAs selecionados, resultando numa listagem final com 18 MBAs. Estes se encontram no topo da Tabela 1, e são seguidos pelos demais MBAs, que não serão detalhados adiante, todos ordenados pelo tamanho da base (o maior valor entre anunciado e estimado).

3.1 Considerações

Algumas considerações sobre os MBAs com base na experiência desta primeira classificação se fazem necessárias:

- **Linguagem:** em alguns MBAs a língua inglesa não é padrão, mas é possível alterá-la já na página inicial;
- **Tipo de registro:** alguns MBAs apontam o total de registros, mas não especificam se são artigos completos, resumos ou citações. Além disso, alguns MBAs declaram possuir uma grande quantidade de registros. Percebeu-se contudo, que grande parte deles não são artigos completos, mas sim resumos e citações que “inflam” o tamanho da base de dados;
- **Replicação:** ao exibir o resultado de uma busca, alguns MBAs podem trazer resultados replicados de uma referência (por exemplo, um resultado mostra o artigo completo, outro apenas o resumo do mesmo artigo e outro apenas a citação), “inflando” o tamanho do resultado;
- **Disponibilização:** para cada registro de documento nos MBAs podem haver três situações: (1) o mecanismo além de fornecer a informação sobre o documento, também mantém e disponibiliza ele em algum servidor; (2) o registro é informativo e indica um link onde o usuário pode encontrar o documento efetivamente; ou (3) o registro é somente informativo, não possui nem link, quanto menos o arquivo. A questão que permeia essa consideração é: um mecanismo que contém somente o registro de um documento possui o mesmo mérito de um mecanismo que possui a informação e é o meio pelo qual se pode acessar o documento? Considera-se que não, pois não seria útil existir vários mecanismos com informações sobre um documento, sendo que este documento não está disponível em lugar algum;
- **Mérito:** percebeu-se que é comum encontrar apenas uma citação/resumo em um MBA que não possui o documento procurado e, com base nas informações encontradas, fazer uma nova busca em outro mecanismo (inclusive nos genéricos) para encontrar o artigo que, por vezes, pode estar hospedado em um terceiro banco de dados. Neste contexto permeiam pelo menos duas questões a serem respondidas: qual MBA tem mérito? Algum deles tem mais mérito que outro? De certa forma são questões relativas às considerações de disponibilização;
- **Tamanho real:** percebeu-se que alguns MBAs anunciam um número de documentos, mas através das buscas genéricas não foi possível encontrar o mesmo valor. Isso pode significar uma informação incorreta ou que

buscas genéricas não são implementadas em alguns MBAs ou ainda que o algoritmo de busca não esteja encontrando todos os registros que poderia encontrar. Pode-se resumir isto como uma questão: O que considerar? Tamanho anunciado vs. tamanho estimado vs. realidade?

O que se conclui desta primeira etapa de seleção dos melhores MBAs é que, mesmo para informações básicas quantitativas, já existe uma dificuldade inerente para comparar os mecanismos entre si.

4 Classificação

Dentre os critérios citados por Oppenheim, Morris e McKnight [14] – ditos como os mais utilizados pelos pesquisadores – alguns já foram utilizados para a primeira filtragem de mecanismos de busca, enquanto que outros são de difícil mensuração e, dentro do contexto desta pesquisa, podem ser duvidosos. O critério de atualidade (relacionado com a corretude dos links, e.g., número de links quebrados) por exemplo, seria de laboriosa mensuração por usuários do sistema, pois é necessário testar uma certa amostra de links associada aos documentos científicos. Dependendo do tamanho da amostra, esse procedimento seria inviável. Por outro lado, o critério de tempo de resposta da ferramenta poderia ser duvidoso em testes realizados por usuários externos ao sistema. Isso porque, quando o mecanismo informa o tempo de busca, não há como garantir a veracidade de tal informação, e quando este tempo não é informado, seria necessário realizar testes considerando vários tipos de instâncias de busca (e.g., usando ou não operadores booleanos, o carácter coringa, frases exatas etc.), além de considerações sobre o tempo consumido em outras etapas, principalmente na camada de rede.

Tentando evitar a subjetividade e efetuar uma comparação objetiva, foram escolhidos os seguintes critérios de Oppenheim, Morris e McKnight [14]: disponibilidade de recursos do sistema e filtros de busca. Além destes critérios, foi considerado que a apresentação de certas métricas relacionadas aos documentos científicos também é um fator que pode caracterizar uma melhor qualidade de um MBA. Considerando estes três critérios objetivos, foram definidos 25 aspectos, agrupados em três macro parâmetros de avaliação:

- Recursos de Busca (RB)

- RB₁ Filtros booleanos (OR, AND, NOT): o MBA permite filtrar com utilização de operadores booleanos na frase de busca?
- RB₂ Filtros de período de publicação: o MBA permite filtrar/restringir um período de publicações?
- RB₃ Carácter coringa: o MBA permite o uso de caracteres coringa (carácter que pode ser trocado por qualquer carácter ou conjunto de caracteres)?
- RB₄ Sinônimos: o MBA inclui automaticamente sinônimos e variações gramaticais das palavras de busca?
- RB₅ Frase exata: o MBA permite consultar com uso de uma frase exata, na qual a sequência de palavras deve estar na mesma ordem em que foi digitada?
- RB₆ Busca por autores: o MBA permite consultar por autores (nome/sobrenome)?
- RB₇ Busca por filiação: o MBA permite consultar por filiação (universidade/laboratório ou grupo de pesquisa)?
- RB₈ Busca por ISBN/ISSN: o MBA permite consultar pelo número de *International Standard Book Number* (ISBN) – para livros – ou *International Standard Serial Number* (ISSN) – para periódicos?
- RB₉ Busca por local de publicação: o MBA permite consultar pela cidade, região ou país onde ocorreu a publicação?
- RB₁₀ Tipo de documento: o MBA permite consultar por um tipo específico de material (e.g. artigo, citação, livro etc.)?
- RB₁₁ Meio de publicação: o MBA permite consultar por artigos publicados em *journals* e conferências, específicas (e.g., Revista Brasileira de Computação Aplicada - RBCA)?
- RB₁₂ Idioma do documento: o MBA permite consultar por artigos escritos em um determinado idioma?

- Recursos de Refinamento (RR)

- RR₁ Filtragem dinâmica: o MBA permite filtrar os resultados de uma busca?
- RR₂ Ordenação por relevância: o MBA permite ordenar documentos por relevância em relação à frase de busca?
- RR₃ Ordenação temporal: o MBA permite ordenar documentos pela data em que foram publicados?
- RR₄ Ordenação por citações: o MBA permite ordenar documentos pelo número de citações contabilizadas?
- RR₅ Revisão: o MBA indica se o documento foi revisado por uma banca avaliadora?
- RR₆ Citações: o MBA indica o número de citações feitas à um determinado documento?
- RR₇ Referências: o MBA indica o número de referências utilizadas no documento?
- Recursos Auxiliares (RA)
 - RA₁ Resumo: o MBA apresenta os resumos dos documentos disponíveis?
 - RA₂ Palavras-chaves: o MBA apresenta o conjunto de palavras-chaves dos documentos disponíveis?
 - RA₃ Lista temporária: o MBA permite a criação de lista temporária de documentos ou gera histórico de consultas?
 - RA₄ Alerta: o MBA permite ativar opção de alerta, avisando sobre atualizações específicas por email?
 - RA₅ Conteúdo relacionado: o MBA sugere conteúdo relacionado à frase de busca ou à um determinado artigo (e.g., resumos e palavras-chave semelhantes ou artigos de um mesmo autor)?
 - RA₆ Ajuda: o MBA possui uma seção de ajuda?

Os 18 MBAs classificados na seleção inicial foram avaliados em relação a cada um dos critérios selecionados. A verificação foi realizada nos dias 25 e 26 de julho de 2013 e os dados coletados foram compilados na Tabela 2, que apresenta os MBAs em ordem alfabética. Cada recurso contemplado por um MBA foi identificado na tabela com o símbolo ●, enquanto que os recursos não contemplados foram identificados com o símbolo ○. Durante a coleta de dados foram anotadas três considerações:

- Alguns MBAs possuíam link de acesso aos documentos, direcionando os usuários a outras bases de arquivos, mas não apresentavam informações de resumo e palavras chaves, por exemplo;
- Alguns MBAs só permitiram a criação de uma lista de artigos selecionados e criação de alertas se o usuário estiver autenticado no sistema. O acesso concedido pela CAPES, contudo, não foi suficiente para realizar este tipo de operação. Todavia, estes mesmos MBAs mantêm registro do histórico de consultas realizadas;
- Nenhum mecanismo informa o número de avaliadores que revisaram os documentos indexados, mas alguns identificam os documentos que foram revisados por pares.

5 Discussão

Considerou-se que existe uma disparidade nos graus de importância entre os grupos de recursos sugeridos. Os recursos de busca (RB) foram considerados os mais importantes, pois englobam as funcionalidades principais de um MBA. Os filtros booleanos, por exemplo, são um dos recursos mais úteis para construção de frases de busca não triviais, possibilitando a realização de buscas complexas e objetivas, gerando um resultado final bem delimitado. Da mesma forma, os recursos de refinamento (RF) agregam maior importância do que os recursos auxiliares (RA), pois podem reduzir o número de documentos falso-positivos (aqueles que apareceram no resultado, mas não tem relação com o que se desejava buscar), ordenar o resultado por um critério que seja do interesse do pesquisador, e apresentar informações significativas sobre documentos.

Baseado nas considerações mencionadas, foi realizada uma ponderação (Equação 1) entre os grupos de parâmetros. Nesta ponderação, a quantidade de RBs (RB_i) identificada como mais importante recebeu o maior peso na equação, sendo multiplicada por três; em sequência, a quantidade de RRs (RR_i) recebeu um peso um pouco menor, sendo multiplicada por dois; e a quantidade de RAs (RA_i) recebeu o menor peso, sendo multiplicada por um. Apesar da ponderação poder ser questionada quanto a importância relativa entre seus critérios, acredita-se que ela é menos prejudicial do que considerar todos os tipos de recurso com a mesma importância.

Tabela 2: Avaliação dos 18 MBAs da seleção inicial quanto à oferta de recursos.

Mecanismos	Recursos de Busca (RB _i)												Recursos de Refinamento (RR _i)							Recursos Auxiliares (RA _i)					
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	1	2	3	4	5	6
ACM DL	●	●	●	○	●	●	●	●	●	●	●	○	●	●	●	●	○	●	●	●	●	○	○	●	○
BASE	●	●	●	●	●	●	○	○	●	●	○	●	●	●	●	○	○	○	○	●	●	●	○	○	●
CCSB	●	●	●	○	●	●	○	○	○	○	○	○	○	●	●	○	○	○	○	●	○	○	○	○	●
CiteSeer ^X	●	●	○	●	●	●	●	○	○	○	○	○	○	●	●	●	○	●	●	●	○	●	○	○	●
DBLP	●	●	●	●	●	●	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○	○	●
Engineering Village	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	○	○	●	●	●	●	○	○	●
Google Scholar	●	●	○	○	●	●	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	○	●	○	○
IEEE Xplore	●	●	●	●	●	●	●	●	○	●	●	○	●	●	●	●	○	●	●	●	●	○	○	●	●
Ingenta Connect	●	○	●	○	●	●	○	●	○	○	○	○	○	●	●	○	○	○	○	●	●	○	○	●	●
inSpire HEP	●	●	●	●	●	●	●	○	○	○	●	○	○	○	●	○	○	●	●	●	●	○	○	○	●
Periódicos da Capes	●	●	●	○	●	●	○	●	○	●	○	●	●	●	●	○	●	○	○	●	●	○	○	○	●
SciELO	●	●	●	○	●	●	○	○	●	○	●	○	●	●	●	○	○	○	●	●	●	●	○	○	○
Science Direct	●	●	●	●	●	●	●	●	○	●	●	○	●	●	●	○	○	○	●	●	●	●	○	●	●
Scirus	●	●	●	○	●	●	●	●	○	●	●	○	●	●	●	○	○	○	○	○	○	●	○	●	●
Scopus SciVerse	●	●	●	○	●	●	●	●	○	●	●	●	●	●	●	●	○	●	●	●	●	●	○	●	●
Socol@r	●	●	●	○	●	●	○	●	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○
Springer Link	●	●	●	●	●	●	○	●	○	●	●	●	●	●	●	○	○	○	●	●	●	○	○	○	●
Web of Knowledge	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	○	●	●	●	●	●	○	●	●

$$\text{Recursos} = 3 \times \sum_{i=1}^{12} \text{RB}_i + 2 \times \sum_{i=1}^7 \text{RR}_i + \sum_{i=1}^6 \text{RA}_i \quad (1)$$

Considerando os resultados obtidos na coleta de dados e a ponderação sugerida, os 18 MBAs analisados nesta segunda etapa foram classificados e apresentados na Tabela 3 em ordem crescente de pontuação de total de recursos. Os resultados encontrados indicam o Web of Knowledge na primeira posição, seguido por Engineering Village na segunda posição, e Scopus SciVerse na terceira posição. A diferença entre os resultados do primeiro e do terceiro MBA foi de apenas três pontos; além disso, os quatro MBAs seguintes na classificação – IEEE Xplore, ACM DL, Science Direct e Springer Link – também ficaram próximos, distantes apenas um, dois, quatro e seis pontos respectivamente em relação ao terceiro MBA classificado. Essa proximidade na classificação mostra que a diferença de recursos entre eles é sutil. Pode-se perceber também que os MBAs mais bem classificados possuem uma alta quantidade de recursos em cada um dos três grupos considerados nesta pesquisa.

Tabela 3: Classificação dos 18 MBAs da seleção inicial quanto à oferta de recursos.

Mecanismos	$3 \times \sum_{i=1}^{12} \text{RB}_i$	$2 \times \sum_{i=1}^7 \text{RR}_i$	$\sum_{i=1}^6 \text{RA}_i$	Recursos
Web of Knowledge	33	12	5	50
Engineering Village	36	8	4	48
Scopus SciVerse	30	12	5	47
IEEE Xplore	30	12	4	46
ACM DL	30	12	3	45
Science Direct	30	8	5	43
Springer Link	30	8	3	41
BASE	27	6	4	37
Scirus	27	6	3	36
Periódicos da Capes	24	8	3	35
inSpire HEP	24	6	3	33
SciElo	21	8	3	32
CiteSeer ^X	18	10	3	31
DBLP	24	2	1	37
Ingenta Connect	15	4	4	23
CCSB	15	4	2	21
Socol@r	18	2	0	20
Google Scholar	12	4	1	17

Calculando a média ($\mu = 35,11$), desvio padrão ($\sigma = 10,13$), e coeficiente de variação ($cv = 29\%$) da pontuação de recursos dos 18 MBAs avaliados, pode-se concluir através de uma análise da distribuição normal que os MBAs classificados nas três primeiras colocações não são muito superiores à média, sendo que o primeiro, segundo e terceiro mecanismos estão 1,47, 1,27 e 1,17 desvio padrão acima da média respectivamente. Contudo, como o coeficiente de assimetria dos valores de recursos é negativo, o que indica maior quantidade de MBAs com pontuação de recursos inferiores à média, pode-se considerar que os seis melhores MBAs são um grupo seletos, pois existem mais MBAs com valores de recursos significativamente inferiores do que MBAs com valores de recursos próximos. O coeficiente de curtose em relação ao resultado dos recursos também foi calculado ($K = 0,33$) e este indicou uma distribuição platicúrtica, i.e. uma distribuição de valores mais achatada do que a distribuição normal.

Com exceções, é possível afirmar que aqueles mecanismos que apresentaram vários recursos de um dos tipos analisados, apresentaram também vários recursos nos outros tipos. Para validar essa hipótese foi realizada uma análise estatística de correlação linear entre os três recursos considerados, comparando-os dois a dois. Foram calculados os coeficientes de correlação (r) e de determinação (r^2) para cada par, sendo: $r = 0,6320$ entre recursos de busca e recursos de refinamento, indicando correlação moderada com $r^2 = 40\%$; $r = 0,6298$ entre recursos de busca e recursos auxiliares, indicando também correlação moderada com $r^2 = 40\%$, e; $r = 0,7119$ entre recursos de refinamento e recursos auxiliares com $r^2 = 51\%$. Esse resultado comprova a hipótese apresentada, mostrando que existe uma correlação moderada entre os recursos busca, refinamento e auxiliares.

Analisando os resultados dos recursos de busca individualmente, foram calculados média, desvio padrão e coeficiente de variação obtendo-se $\mu = 24,67$, $\sigma = 6,67$ e $cv = 0,27$ respectivamente. Verificou-se que apenas o MBA Engineering Village obteve a maior pontuação, apresentando todos os recursos elencados. Por outro lado, o MBA Google Scholar apresentou o pior resultado, possuindo menos da metade dos recursos elencados (quatro de doze). Esse resultado reforça os achados de Jacso [8], há oito anos, quando criticou o Google Scholar por não possibilitar a realização de buscas avançadas. O recurso de busca por local de publicação (RB₉) foi o menos encontrado, sendo que apenas 4 MBAs apresentaram tal recurso.

Considerando os resultados dos recursos de refinamento, foram calculados média, desvio padrão e coeficiente de variação obtendo-se $\mu = 7,33$, $\sigma = 3,27$ e $cv = 0,45$ respectivamente. Nenhum MBA apresentou todos os recursos de refinamento elencados, contudo 4 deles (22%) apresentaram seis dos sete recursos. Por outro lado, 5 MBAs (28%) apresentaram menos do que três recursos de refinamento. O recurso de refinamento menos encontrado foi o de revisão (RR₅), encontrado em apenas 2 MBAs (11%). A operação de ordenação por número de citações (RR₄) também foi um recurso pouco encontrado (apenas em 5 dos 18 MBAs – 28%).

Com relação aos resultados sobre os recursos auxiliares, foram calculados média, desvio padrão e coeficiente de variação obtendo-se $\mu = 3,11$, $\sigma = 1,37$ e $cv = 0,44$ respectivamente. Nenhum MBA possuía todos os recursos auxiliares, mas 3 deles (17%) apresentaram cinco dos seis recursos analisados. O Socol@r foi o único a apresentar nenhum dos recursos auxiliares analisados, enquanto que o Google Scholar foi o único a disponibilizar gratuitamente o recurso de alerta, sendo que alguns outros MBAs (ACM DL, Engineering Village, IEEE Xplore, Scopus, Web of Knowledge) também permitem tal operação, mas somente com acesso mais amplo do que o concedido pela CAPES.

6 Conclusão

Com a vasta disponibilidade de informações científicas na Web, torna-se importante identificar bons mecanismos de busca acadêmica (MBA), bem como saber utilizá-los. Este artigo analisou 40 MBAs em duas fases de classificação: uma macro, eliminatória, identificando características gerais; e outra micro, classificatória, identificando recursos do mecanismo para os momentos de busca. Apenas 18 mecanismos foram selecionados pela primeira fase e depois classificados. Os MBAs que se apresentaram nas sete primeiras posições da classificação, com pontuação geral acima de meio desvio padrão da média ($\mu + 0,5\sigma$), foram respectivamente: Web of Knowledge, Engineering Village, Scopus SciVerse, IEEE Xplore, ACM DL, Science Direct e Springer Link.

Para se chegar a este resultado, observou-se uma dificuldade na obtenção e confiabilidade de dados, bem como na definição e seleção de critérios comparativos, além do fato de que novos MBAs podem estar surgindo e da listagem inicial não incluir todos os mecanismos existentes. Deve-se ter ciência também de que a evolução dos mecanismos é abrupta e, portanto, esta pesquisa apresenta resultados para o contexto atual. Entretanto, acredita-se que os MBAs mais bem classificados representam certamente um bom conjunto de sites para se iniciar qualquer pesquisa na área tecnológica.

Independentemente da análise ter sido realizada sobre a forma ou conteúdo, em ambos os casos sempre há um fator pessoal muito grande que influi na resposta da pesquisa através de MBAs, que é a sensibilidade e o conhecimento do pesquisador para ajustar os parâmetros de busca – tais como palavras-chave e refinamentos – para realmente conseguir extrair os artigos interessantes para sua temática de pesquisa. Reconhece-se que dois outros conjuntos de fatores subjetivos podem influenciar o resultado do uso dos mecanismos de busca científicos durante uma pesquisa bibliográfica: a usabilidade da interface e a preferência (associada à experiência) do pesquisador no uso e conduta da pesquisa. Estas análises, portanto, podem ser executadas como trabalhos futuros.

Por fim, ousamos recomendar aos pesquisadores que, além de usar os melhores MBAs que possam ter acesso, executem suas pesquisas sempre em mais de um deles e procurem utilizá-los além dos recursos triviais de busca sob pena de estarem sendo pouco produtivos nesta fase de suas pesquisas ou não estarem obtendo as melhores referências para suas pesquisas.

7 Agradecimentos

Agradecemos a todos os discentes do Mestrado Acadêmico em Computação Aplicada da Universidade do Estado de Santa Catarina pelas discussões e sugestões a esta pesquisa. Agradecemos também pelas bolsas de mestrado concedidas pela Universidade do Estado de Santa Catarina através do programa PROMOP.

Referências

- [1] MOSTAFA, J. Seeking Better Web Searches. *Scientific American*, 292(2): 51–57, 2005.
- [2] ATSAROS, G.; SPINELLIS, D.; LOURIDAS, P. Site-specific Versus General Purpose Web Search Engines: A Comparative Evaluation. In: *IEEE Conference on Panhellenic PCI'08*, Samos, 2008. p. 44–48.
- [3] BEEL, J.; GIPP, B.; WILDE, E. Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar & Co. *Journal of Scholarly Publishing*, 41(2): 176-190, 2010.
- [4] RIEGER, O. Y. Search Engine Use Behavior of Students and Faculty: User Perceptions and Implications for Future Research. *Peer-Reviewed Journal on the Internet*, 14(12), 2009. Disponível em: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2716/2385>>. Acesso em: 05 ago. 2013.
- [5] BROPHY, J.; BAWDEN, D. Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources. *Aslib Proceedings*, 57(6): 498-512, 2005.
- [6] MEIER, J. J.; CONKLING, T. W. Google Scholar's Coverage of the Engineering Literature: An Empirical Study. *Journal of Academic Librarianship*, 34(3):196–201, 2008.
- [7] MEHO, L. I.; YANG, K. Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13):2105–2125, 2007.
- [8] JACSO, P. As we may Search - Comparison of Major Features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9):1537-1547, 2005.
- [9] FALAGAS, M.; PITSOUNI, E.; MALIETZIS, G.; PAPPAS, G. Comparison of Pubmed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses. *The FASEB Journal*, 22(2):338–342, 2008.
- [10] TURNER, M. Digital Libraries and Search Engines for Software Engineering Research: An Overview. Technical Report, 2010. Disponível em: <http://www.dur.ac.uk/ebse/resources/notes/tools/SearchEngineIndex_v5.pdf>. Acesso em: 05 ago. 2013.
- [11] JANSEN, B.; SPINK, A. How are we Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. In *Information Processing and Management*, v.42, 2006. p. 248–263.
- [12] EARNEY, L. An Introduction to the JISC Academic Database Assessment Tool. *Serials: The Journal for the Serials Community*, 22(3):199–205, 2009.
- [13] ZHAO, D. Z.; LOGAN, E. Citation Analysis Using Scientific Publications on the Web as Data Source: A Case Study in the XML Research Area. *Scientometrics*, 54(3), 449-472, 2002.
- [14] OPPENHEIM, C.; MORRIS, A.; MCKNIGHT, C.; LOWLEY, S. The Evaluation of WWW Search Engines. *Journal of Documentation*, 56(2): 190-211, 2000.

Apêndice

Tabela 4: Lista dos Mecanismos de Busca Acadêmicos Considerados.

Mecanismos de Busca Acadêmica	Web Site
#01 [Aca. Pub. eJournal]	http://apejournal.weebly.com/
#02 [ACM DL]	http://dl.acm.org/
#03 [Airiti]	http://portal.airiti.com
#04 [AJO]	http://www.ajol.info/
#05 [ArnetMiner]	http://arnetminer.org/
#06 [arXiv]	http://arxiv.org/
#07 [BASE]	http://www.base-search.net/
#08 [CCSB]	http://liinwww.ira.uka.de/bibliography/
#09 [CiteSeer ^X]	http://citeseerx.ist.psu.edu/
#10 [Citeulike]	http://www.citeulike.org/
#11 [Cogprints]	http://cogprints.org/
#12 [DBLP]	http://www.informatik.uni-trier.de/~ley/db/
#13 [Emerald]	http://www.emeraldinsight.com/
#14 [Engineering Village]	http://www.engineeringvillage2.org/
#15 [GJS]	http://journalseek.net/
#16 [Google Scholar]	http://scholar.google.com.br/
#17 [IEEE Xplore]	http://ieeexplore.ieee.org/
#18 [ICI]	http://www.indiancitationindex.com/
#19 [Ingenta] Connect	http://www.ingentaconnect.com/
#20 [inSpire] HEP	http://inspirehep.net/
#21 [Mendeley]	http://www.mendeley.com/
#22 [Microsoft AS]	http://academic.research.microsoft.com/
#23 [NDL]	http://www.ndl.go.jp/en/
#24 [NSDL]	http://nsdl.org/
#25 [OCLC]	http://www.oclc.org/oaister/
#26 [Periódicos da Capes]	http://periodicos.capes.gov.br/
#27 [Pubget]	http://pubget.com/
#28 [SafetyLitv]	http://www.safetylit.org/
#29 [SciFinder]	http://www.cas.org/products/scifinder
#30 [SciDiver]	http://www.scidiver.com/
#31 [SciELO]	http://www.scielo.org/
#32 [Science Accelerator]	http://www.scienceaccelerator.gov/
#33 [Science.gov]	http://www.science.gov/
#34 [Science Direct] SciVerse	http://www.sciencedirect.com/
#35 [Scirus]	http://www.scirus.com/
#36 [Scopus] SciVerse	http://www.scopus.com/
#37 [Springer Link]	http://www.springerlink.com/
#38 [Socol@r]	http://www.socular.com/
#39 [Web of Knowledge]	http://apps.webofknowledge.com/
#40 [WorldWideScience.org]	http://worldwidescience.org/