

Desenvolvimento de uma nova Ferramenta Computacional para Auxílio na Pesquisa Acadêmica Avançada

Kleber Zuza Nobrega¹, Edeilson Pereira Pestana²

¹Departamento de eletroeletrônica – Instituto federal de educação, ciência e tecnologia do Maranhão (IFMA)

Av. Getúlio Vargas, 4 -Monte Castelo, São Luis – MA, CEP 65030-005

²Departamento de eletroeletrônica – Instituto federal de educação, ciência e tecnologia do Maranhão (IFMA)

Av. Getúlio Vargas, 4 -Monte Castelo, São Luis – MA, CEP 65030-005

bzuzal@yahoo.com.br, edeilsonpestana@gmail.com

Abstract. *The academic research still finds barriers, one of them is the difficulty of doing research for papers accurately, whether the inaccuracy of current search engines or the fact the some are not available free of charge, then comes the need for a system to do the research for scientific production in an efficient, reliable and free. This paper proposes an innovator system, developed in MATLAB, for web search of scientific papers. The search engine, here presented, works with pattern recognition in organizations websites of international academic prestige publishers and returns to the user scientific work obeying search parameters that can be configured for an optimized search.*

Resumo. *A pesquisa acadêmica ainda encontra barreiras, uma delas é a dificuldade de se fazer pesquisa por artigos de forma precisa, seja pela imprecisão dos atuais motores de busca ou pela não gratuidade, surge então a necessidade de um sistema que faça essa pesquisa por produção científica de maneira eficiente, confiável e gratuita. Neste trabalho é proposto um sistema inovador, desenvolvido em MATLAB, para a busca web de trabalhos científicos. O motor de busca aqui apresentado, trabalha com reconhecimento de padrões em sites de organizações e editoras de prestígio acadêmico internacional e retorna ao usuário trabalhos científicos obedecendo parâmetros de busca que podem ser configurados para uma busca otimizada.*

1. Introdução

A produção científica global vem crescendo ano após ano e grande parte desse aumento da produtividade se deve ao desenvolvimento tecnológico com o advento da internet que vem auxiliando pesquisadores à realizarem estudos se valendo de técnicas e ferramentas que só são possíveis pela facilidade que tal desenvolvimento vem proporcionando ao homem.

Uma das principais características da pesquisa científica é a necessidade da compreensão total de um determinado tema que esteja em questão para explicar a realidade como afirma Schwartzman (2002), e isso só é possível quando se tem acesso à todas as

informações referentes ao tema que se pretende estudar, daí parte a necessidade de ser ter acesso de forma objetiva à estudos de outros pesquisadores para se obter dados sobre o estado da arte, segundo Salem et al (2009). O problema é que o acesso à tais estudos nem sempre se dá de forma simples, eficaz ou gratuita. A busca por artigos científicos, muitas vezes, é dispendiosa de tempo já que os motores de busca atuais não garantem uma busca refinada e precisa.

O *google* acadêmico ou *scholargoogle*, por exemplo, é uma ferramenta de pesquisa que permite a busca por trabalhos acadêmicos, artigos, jornais de universidades entre outros. Todavia, pesquisadores experientes conhecem a necessidade da pesquisa que obedeça a parâmetros de busca bem definidos, algo que o *scholargoogle* não oferece. É possível pesquisar determinado tema, autor e período específico, no entanto essa busca retorna resultados que muitas vezes são dispensáveis ao pesquisador experiente.

Outra alternativa para a pesquisa de artigos e estudos em geral é a plataforma *webofScience*, que assim como o *scholargoogle* é uma ferramenta muito poderosa como motor de busca, mas essas buscas nem sempre retornam resultados refinados de determinado autor, específicos periódicos e, além disso, é uma ferramenta paga, o que não a torna popular por discentes, centros de pesquisa, universidades e faculdades.

A proposta deste trabalho é uma ferramenta de busca científica, inicialmente, na área da física, devido à demanda de alguns pesquisadores nesta área, desenvolvida em MATLAB que permite à pesquisadores efetuarem buscas por artigos nos mais renomados periódicos internacionais usando, especificamente, a qualificação *qualis* da CAPES, que é referência utilizada por pesquisadores brasileiros de alto padrão acadêmico.

2. Justificativa

A principal motivação para a realização desde trabalho é justamente a inexistência de uma ferramenta que possa fazer a busca especializada e com a interface simples da forma que este trabalho propõe, e que esteja à disposição dos pesquisadores mais experientes e ao alcance daqueles que desejam fazer pesquisa de alto padrão, seja acompanhando determinado tema ou mesmo acompanhamento a produção de determinado pesquisador.

O trabalho consiste em um programa de computador, desenvolvido em MATLAB, por este ser um *software* muito utilizado nas universidades em todo o mundo e possuir uma grande variedade de aplicações, onde é possível efetuar a busca de forma automática de artigos científicos nos sites da *OSA*, *APS* e *ELSEVIER*, onde o usuário do programa tem a possibilidade de determinar parâmetros de busca como:

- a) Busca por autor;
- b) Busca por tema;
- c) Busca por período específico;
- d) Busca em determinado periódico e revista;

Definidos os parâmetros, o programa efetua a busca nos sites dos periódicos e retorna:

- 1) Os títulos dos artigos;
- 2) Os nomes dos autores;
- 3) Os *links* que direcionam à página contendo o abstract dos artigos;

- 4) Os *links* dos arquivos PDF dos artigos;
- 5) O texto do resumo dos artigos;
- 6) Os dados dos artigos (nome da revista, volume, páginas);

Além disso, é importante ressaltar que a ferramenta aqui desenvolvida é um programa executável e possui uma interface gráfica (GUI) bastante amigável, além disso o usuário não necessita ter MATLAB instalado em seu computador, e nem acesso às plataformas pagas pela CAPES, a única necessidade é o acesso à internet.

Nas seções seguintes, estar-se-á explicitado os métodos utilizados no desenvolvimento do algoritmo, e toda a dinâmica de busca, bem como os materiais que foram utilizados neste trabalho, além de demonstrar, através de exemplo, como o programa pode ser utilizado para a busca de artigos científicos.

3. Material e métodos

Os requisitos mínimos para execução da ferramenta são: processador x86-64 AMD ou Intel; sistema operacional Windows 7, 8, 8.1 ou 10; Memória RAM de 2GB e 2GB de espaço em disco.

O algoritmo e a estrutura do programa foram escritos fazendo-se uso de várias funções do MATLAB onde se destacam as mais relevantes para o desenvolvimento da ferramenta como sendo a função *webread()*, capaz de fazer leitura de conteúdo de sites da *web* (ou URL); a função *regexp()*, usado para fazer reconhecimento de padrões em *strings*, além da capacidade do *software* em lidar com diferentes tipos de variáveis; e por fim, a fácil implementação e integração de todas as ferramentas em uma única interface gráfica, e do tipo .exe.

A função *webread()* funciona de tal forma que o parâmetro de entrada da função neste trabalho é o *link* da página que se deseja ler (*webread(URL)*) e então a função retorna o código fonte desta contendo *hiperlinks*, textos formatados em HTML.

Uma vez obtida a informação da função *webread*, a função *regexp()* que utiliza como parâmetro de entrada uma *string* e um padrão programável a depender do interesse do usuário, será utilizada para efetuar a busca pelo padrão que se deseja reconhecer e um *token* nomeado, que nada mais é que o objeto que foi encontrado (pode ser um nome, uma data, um número, um link). Como no exemplo didático da figura 1 abaixo.

```
%% exemplo
fonte = '2016< titulo "estudo da física" >2016';
padrao = '2016< titulo "(?<TITULO>.*?)" >2016';

padrao_encontrado = regexp(fonte,padrao,'names')
```

Figure 1- exemplo do uso da função *regexp()*.

Neste exemplo a variável denominada de fonte é o arquivo de texto onde deveremos fazer a busca por padrões, já o padrão representado é a informação que se deseja obter do texto fonte, nesse caso a frase “estudo da física” foi substituída por um token que será o que se deseja capturar do texto fonte. A figura 2 exemplifica o resultado da execução dessa função para este caso.

```
padrao_encontrado =
    TITULO: 'estudo da física'
```

Figure 2 - resultado do código da figura 1.

Com base no exemplo dá para se ter uma ideia das possibilidades que esta função oferece já que a identificação de padrões é crucial à programas de busca. Utilizando esta ferramenta podemos obter links, *strings*, datas entre outras informações presentes nos códigos fonte dos sites da *OSA*, *Elsevir* e *APS*.

3.1. Algoritmo

Inicialmente o usuário define os parâmetros de busca (revista, autor, tema e período). Feita a parametrização, o programa carrega a página principal dos periódicos. Como pode ser visto na figura 3 ilustra a página inicial na qual é efetuada uma busca para se obter os nomes das periódicos e os respectivos links para acesso para a *OSA*.



Figure 3 - Página da OSA com todas as revistas (na figura apenas duas).

O algoritmo limita sua busca somente àqueles periódicos que foram selecionadas pelo usuário da plataforma. Depois de feita a eleição das revistas selecionadas, o algoritmo faz o carregamento da página de cada um dos periódicos bem como seleciona os volumes e edições para que finalmente passo a ser feita a busca, no código fonte do site, por todos os links que darão acesso a determinada edição da revista. Ver Figura 4.



Figure 4 - Página de um periódico (links das edições e datas em destaque)

Obtidos os *links* das edições e suas datas, o algoritmo faz a validação temporal como base no período especificado pelo usuário, ou seja, o algoritmo seleciona para a busca somente as edições que estão dentro do período em que o usuário deseja efetuar a busca dos artigos.

Seguindo a ordem de processos o algoritmo carrega os *links* das edições válidas para a busca e obtém os títulos de todos os artigos das mesmas, bem como o *link* para acesso ao PDF, os nomes dos autores, o *link* para acesso ao texto do *abstract* e os dados de referência do artigo, como volume e número da edição. Esses dados podem ser vistos na Figura 5.



Figure 5- Página de artigos de determinada revista.

Em posse de todas as informações faz-se, nesse momento, a seleção dos artigos baseados nas escolhas por tema ou autor já definido pelo usuário no início. O algoritmo verifica se algum dos artigos encontrados corresponde aos desejados pelo usuário.

O algoritmo verifica os parâmetros que foram selecionados pelo usuário e então retorna, os resultados da busca daqueles artigos que satisfazem estes parâmetros, ou seja, o resultado entregue ao usuário é em arquivo no formato HTML contendo uma lista com os dados dos artigos que satisfizeram os parâmetros da busca, i.e.; título, *link* do *abstract*, *link* do artigo em formato PDF, dados de referências (volume, periódico, edição, páginas) onde esses dados podem ser acessados gratuitamente.

A escolha do formato HTML para apresentação dos resultados foi feita pensando na facilidade de acesso aos *links* gerados e a possibilidade de o arquivo da busca ser mantido para consulta posterior observando que o público alvo desta ferramenta são, no geral, estudantes de pós-graduação e doutores pesquisadores que realizam revisões bibliográficas quase que de forma constante e necessitam ter tais informações ao alcance.

No fluxograma da figura 6 abaixo é possível ter uma idéia geral dos processos que ocorrem no algoritmo.

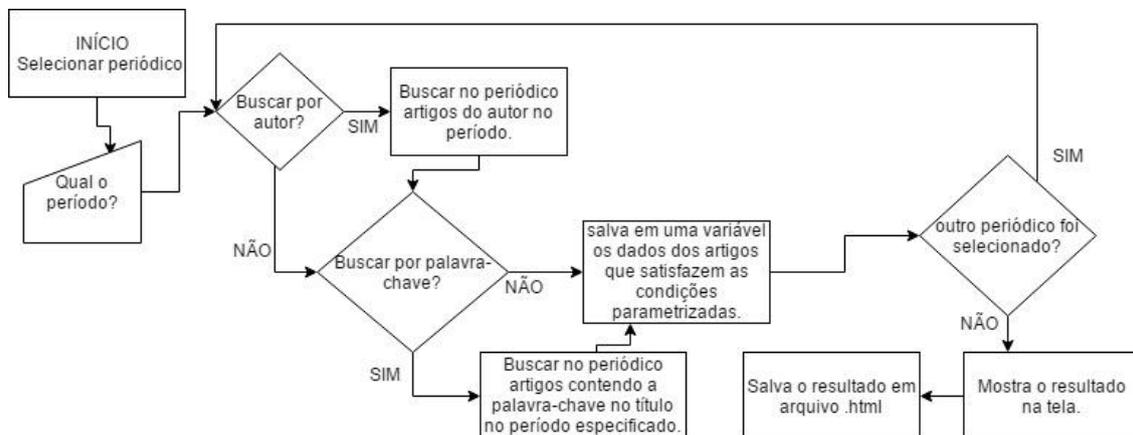


Figure 6 - fluxograma de visão geral do processo de busca.

4. Uso da ferramenta

Aqui será feita uma prova de conceito capaz de ilustrar e exemplificar como o usuário pode fazer uso da ferramenta para a pesquisa de artigos. Neste exemplo a busca será feita por autor, por palavra que contenha no título, que nós denominamos aqui como tema e em período especificado.

4.1. Demonstração do uso da ferramenta

A seguir estudaremos uma busca por uma palavra chave que esteja no título do artigo e pelo nome de autor em período específico. O código pode ser visto na Figura 7 com resultados da busca na Figura 8.

1. Nome do autor: Zhaotao Zhou
2. Palavra chave no título: *photovoltaic*
3. Período: nos últimos 4 meses
4. Periódicos: Optics Express

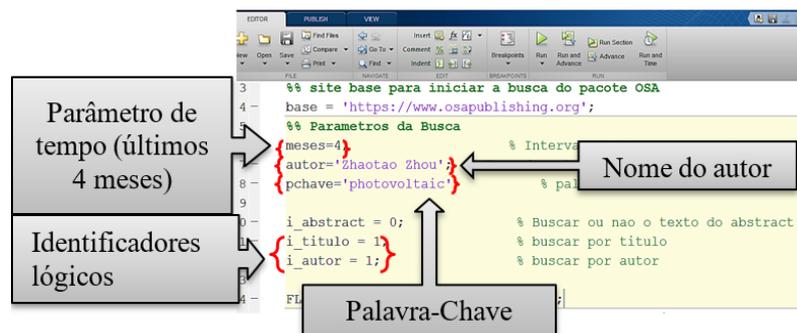


Figure 7- Código no Matlab para buscar por autor, palavra-chave em período definido.

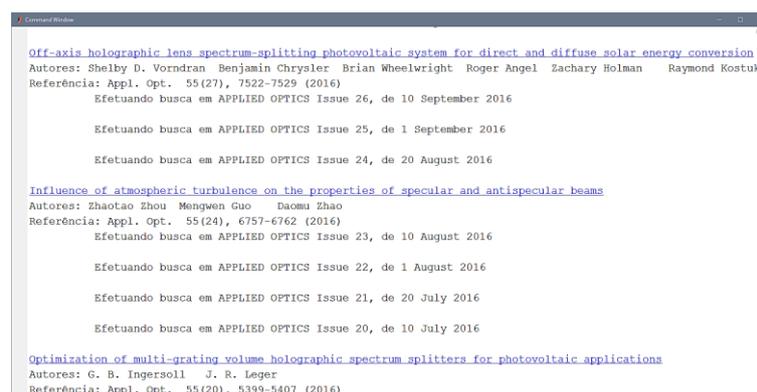


Figure 7 - Resultado da busca por autor, palavra-chave em período e periódico específico com os links para os respectivos resultados.

Com esta demonstração foi possível verificar que os resultados obtidos obedecem rigorosamente os parâmetros de busca, ou seja, todos os artigos que retornam ao usuário contêm a palavra-chave escolhida ou é produto do autor selecionado para a busca e

todos estão dentro do período estipulado (nos últimos 4 meses) além de todos pertencerem ao periódico escolhido para esta busca.

4.2. Interface gráfica

A Interface gráfica, que pode ser vista na figura 9, foi construída pensando em facilitar o seu uso para que mesmo com escolhas simples de parâmetros os resultados sejam estritamente selecionados obedecendo rigidamente os parâmetros desejados.

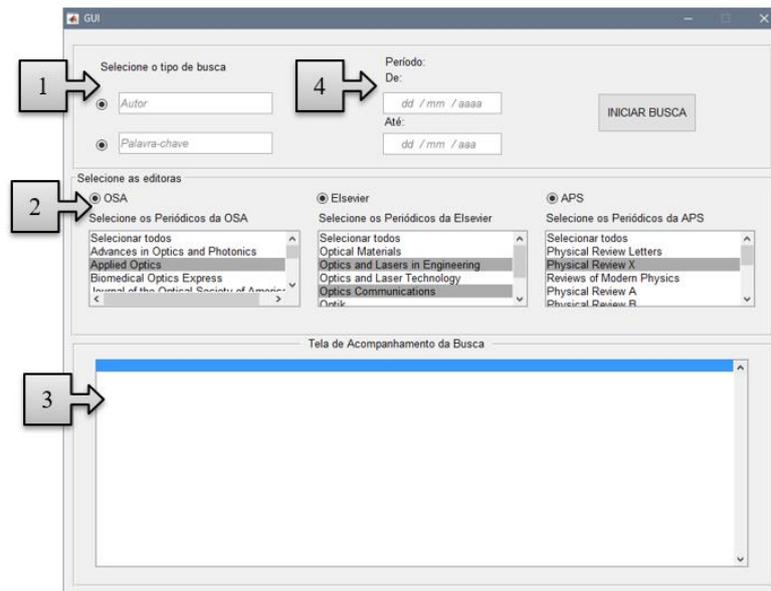


Figure 8 - Interface Gráfica do programa de busca.

Onde as partes enumeradas da figura 8 são:

1. Escolha do autor e palavra-chave
2. Seleção das editoras e respectivos periódicos
3. Tela de acompanhamento de resultados
4. Seleção do período para busca

A escolha por tipo de busca e editoras é facilitada por botões onde o usuário somente necessita selecionar o botão e digitar o nome do autor por exemplo e no caso da seleção de editoras, abrem-se janelas com os nomes dos periódicos bastando o usuário escolher os periódicos que deseja efetuar. A Tela de acompanhamento serve para a visualização do processo de busca em tempo real.

5. Resultados e discussões

O algoritmo desenvolvido em MATLAB tem a grande vantagem de usar linguagem de programação de alto nível dispendo de ferramentas e funções que facilitam muito a programação ampliando as possibilidades de aplicações dos programas.

Os estudos de casos aqui apresentados reforçam a eficiência da ferramenta na busca refinada e precisa por artigos científicos não apresentando resultados que fujam aos parâmetros de busca definidos.

Apesar de a ferramenta aqui apresentada estar definida para busca de artigos na área da física, ela pode ser ampliado para abranger diversas outras áreas do conhecimento e editoras (SPRINGER, IEEE, IOP etc.), tendo em vista que os sites onde as buscas são efetuadas obedecem à regras de padronização inerentes às linguagens nas quais são construídos (HTML e *JAVASCRIPT*), padrões esses fundamentais para a construção do algoritmo em MATLAB, além da possibilidade do desenvolvimento de um aplicativo para dispositivos móveis se valendo da proposta do programa de busca.

6. Conclusão

Os resultados mostram, como esperado, que o sistema é bastante eficiente na busca por artigo, levando em consideração os parâmetros de busca definidos que garantem uma busca refinada e precisa nos sites das organizações e editoras que este trabalho abrange.

A possibilidade de combinar parâmetros é uma das maiores vantagens já que dessa forma os resultados podem variar de acordo com a necessidade do pesquisador que pode particularizar a busca para focar no que realmente importa, que é a pesquisa acadêmica de alto nível.

7. Agradecimentos

Os autores agradecem ao PIBITI e ao Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA) pelo apoio dado ao desenvolvimento da pesquisa.

Referências

Mathworks, matlab. Vs. R2015a, 2015. The MathWorks, Inc. 1 CD-ROM.

Salem, S. and Kawamura, M.D., (2009). Estado da arte dos estados da arte da pesquisa em ensino de física. In: Encontro nacional de pesquisa em ensino de ciências. Ed. VII. Disponível em: <http://posgrad.fae.ufmg.br/posgrad/viiienpec/pdfs/novo_06.pdf>, acesso em 18 set. 2016

Schwartzman, S. (2002). A pesquisa científica e o interesse público. Revista Brasileira de Inovação, São Paulo. Vol. 1, nº 2. Disponível em: <<http://ocs.ige.unicamp.br/ojs/index.php/rbi/article/view/248>>. Acesso em: 19 set. 2016