

UMA PROPOSTA PARA O GERENCIAMENTO *ONLINE* DE INFORMAÇÕES HETEROGÊNEAS DE NATUREZA BIOTECNOLÓGICA

Oliveira G.L.¹, Tamariz A.R.², Gonzalez S. M.³

¹ UENF/Laboratório de Ciências Matemáticas, gl.oliveira@gcomp.uenf.br

² UENF/Laboratório de Ciências Matemáticas, annabell@uenf.br

³ UENF/Laboratório de Ciências Matemáticas, sahudy@uenf.br

Resumo – O NCBI (*National Center for Biotechnology Information*) é um sistema de informações sobre Biologia molecular. Possui como uma de suas funcionalidades a busca por artigos de caráter biológico, através do PubMed. Este oferece poucas e pobres funcionalidades ao usuário na sua interface, pois a obtenção dos resumos dos artigos não é fácil e uma simples busca pode abrir múltiplas janelas. Com a grande quantidade de dados disponível, torna-se muito difícil encontrar artigos relevantes para uma determinada busca sem uma interface apropriada. Este trabalho tem como objetivo oferecer ao usuário um sistema, o *NCBI ESearch Viewer*, com uma interface aprimorada de busca ao portal do NCBI e da base PubMed. Dentre as funcionalidades do sistema encontram-se: fácil acesso às informações dos artigos, navegação por abas nas buscas, filtragem por palavras-chave e por relevância (fator de impacto) do jornal do artigo.

Palavras-chave: Interface Web, Ferramenta de busca, NCBI, PubMed.

Área do Conhecimento: Bioinformática.

Introdução

O Portal NCBI [1] possui um banco de dados denominado PubMed [2], onde constam cerca de 19 milhões de artigos científicos de caráter biológico, além de sequências de proteínas, genomas e outros. Este oferece poucas e pobres funcionalidades ao usuário na sua interface, pois a obtenção dos resumos dos artigos não é fácil e uma simples busca pode abrir múltiplas janelas. Com a grande quantidade de dados disponível, torna-se muito difícil encontrar artigos relevantes para uma determinada busca sem uma interface apropriada. Este trabalho tem como objetivo oferecer ao usuário um sistema, o *NCBI ESearch Viewer*, com uma interface aprimorada de busca ao portal do NCBI e da base PubMed. Junto a um pesquisador da área e usuário frequente do portal, levantou-se que seria interessante, para auxiliar a pesquisa do

usuário, criar uma interface de sistema com recursos para filtrar os resultados obtidos por palavras-chave. Além disso, foi proposto, construir um recurso para filtragem por Índice de Impacto dos artigos.

Metodologia

A Figura 1 mostra o fluxo de atividade desde a pesquisa do usuário até a visualização na interface.

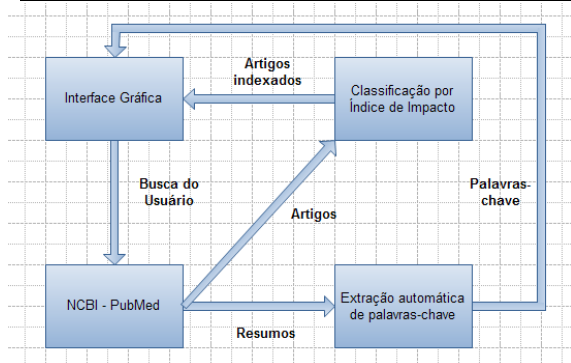


Figura 1- Visão geral do processamento do NCBI ESearch Viewer.

A seguir são apresentadas as funcionalidades do sistema *NCBI ESearch Viewer*. O sistema classifica cada artigo quanto ao seu Índice de Impacto da publicação, e disponibiliza palavras-chave em inglês, extraídas automaticamente, a partir dos resumos de artigos retornados de uma pesquisa feita ao PubMed, que descrevem os assuntos abordados nos artigos. Assim, o usuário pode: filtrar os artigos pelo seu Índice de Impacto; por palavra-chave; visualizar uma *tooltip* (Figura 2), que informa o título do artigo, autores, resumo, palavras-chave e ano de publicação do artigo selecionado. Além disso, cada busca ou filtro realizado é exibido numa nova aba dentro do sistema.

Additional Information:

Journal Impact Factor: 1,81

d.o.i: 10.1267/science.040579197

Title: Nutritional carcinogenesis.
Sutandjo, Noorwati

Abstract: Human beings are often being exposed to carcinogenic factors during their life, some of which are the nutritional factors. From the mechanistic view, nutritional factors are classified into genotoxic and non-genotoxic agents. Genotoxic agent begins their action at the DNA level, causing DNA damage through several mechanisms, e.g. gene point mutations, deletions and insertions, recombinations, rearrangements and amplifications, as well as chromosomal aberrations. Most genotoxic agents are micro components of nutrition, i.e. polycyclic aromatic hydrocarbon (PAH) or heterocyclic amines (HCAs), aflatoxin, and N-nitrosamine. Non-genotoxic agents are less defined in their modes of action, but they are presumed to indirectly affect the cell through tumor promoters. These agents are generally macro components, e.g. high fat. Moreover, epigenetic factors, including changes in the DNA methylation pattern, and peroxidation process resulting reactive oxygen species (ROS), are also known to cause cancer. On the other hand, it is also well recognized that diet and nutrition contain components that can reduce the risk of cancer, in some cases by decreasing the effects of food mutagens, or through carcinogen detoxification, or protection of DNA from electrophilic carcinogen. Thus nutritionally related cancer ultimately develops from an imbalance of carcinogenesis and anti-carcinogenesis process.

Keywords: Humans; Models, Theoretical; Nursing Administration Research; Nursing Staff; Personnel Staffing and Scheduling;
Year: 2010

Figura 2 – Tooltip ampliada

A busca por artigos pode ser filtrada pelo Índice de Impacto usando a seguinte classificação:

- **A1** para artigos com Índice de Impacto igual ou superior a 5.4;
- **A2** para valores menores que 5.4 e maiores ou iguais a 3.2;
- **B1** para valores menores que 3.2 e maiores ou iguais a 2.3;
- **B2** para valores menores que 2.3 e maiores ou iguais a 1.6;
- **C** para valores menores que 1.6.

Para facilitar o reconhecimento das abas manipuladas pelo usuário, o nome da aba é identificado pelo caminho que levou até ela. Por exemplo, como mostra a Figura 3, a terceira aba é o resultado de um filtro composto pela palavra-chave 'Humans', a partir da aba 'human beings'.

Search terms: human beings Search Number of articles: 50 New Search

virus human beings human beings(AZ) human beings: Humans

Close

Filter: All A1 A2 B1 B2 C

Palavras-chave: Nursing Administration Research Nursing Staff Altruism Faculty, Medical History, 20th Century ; Faculty, Medical

PubMed's Article List

Title	Year
Genetics of human social behavior.	2010
Nutritional carcinogenesis.	2010

Figura 3 – Resultado de uma busca.

Para o desenvolvimento da interface, foi preciso escolher ferramentas que possibilitassem interfaces ricas e assíncronas. Portanto, foram utilizados JSF [3] e IceFaces [4], feitas em Java. As requisições ao PubMed utilizaram a plataforma EUtils [5], disponibilizada pelo NCBI. A comunicação utiliza o padrão XML [6].

Resultados

O modo assíncrono de funcionamento do serviço permite que várias buscas sejam efetuadas em paralelo. Um exemplo pode ser

visto na Figura 3 (primeira e segunda aba), onde foi realizada a busca pelo termo ‘*human beings*’, já tendo sido realizada uma busca pelo termo ‘*virus*’.

As abas possibilitam a filtragem de resultados por palavra-chave, como mostra a Figura 3 (quarta aba). A busca por palavras-chave nos resumos dos artigos, hoje, contabiliza apenas palavras-chave previamente definidas pelos autores, visualizando as cinco mais frequentes. Futuramente, vai ser utilizada uma ferramenta para extração automática de palavras-chave, que está em desenvolvimento por outro integrante deste grupo de pesquisa.

O método de exibição de *tooltip* funcionou adequadamente, como pode ser visto na Figura 2, exibindo a informação de um dos artigos listados na Figura 3.

O sistema de reconhecimento do Índice de Impacto do artigo mostrou-se funcional. A terceira aba da Figura 3 é o resultado de um filtro **A2** nos Índices de Impacto da aba ‘*human beings*’. Alguns artigos não conseguiram ser classificados, pois, em alguns casos, o ISSN das publicações está incompleto na fonte de referência utilizada.

O tempo de resposta desde a submissão da requisição, passando pela comunicação com o Eutils para realização das buscas, até a visualização dos resultados vai ser avaliado quando o sistema for implantado e disponibilizado aos pesquisadores num servidor de aplicações.

Discussão

Ao testar o *NCBI ESearch Viewer*, notou-se que algumas funcionalidades precisam ser melhoradas: acessos por múltiplos usuários; o método para atualizar o Índice de Impacto dos jornais; a capacidade de cadastro de usuário no sistema, para que este possa ser personalizado. O cadastro possibilitará configurar a interface de acordo com escolhas do usuário quanto a: salvar as buscas; número de palavras-chave

visualizadas em cada aba; método de ordenação dos resultados (hoje é feita começando pelo artigo mais recente); dentre outras.

Conclusão

O sistema desenvolvido está caminhando para cumprir o proposto. O *NCBI ESearch Viewer* já possui buscas assíncronas, exibição de *tooltip* para fácil acesso a informações dos artigos, filtro por Índice de Impacto, navegação por abas e filtro por palavras-chave, podendo ser colocado em fase de avaliação.

Referências

- [1] U.S. National Library of Medicine. National Center for Biotechnology Information. Disponível em: <<http://www.ncbi.nlm.nih.gov/guide>>. Data de acesso: 23/04/2010.
- [2] U.S. National Library of Medicine. Pubmed. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed>>. Data de acesso: 23/04/2010.
- [3] Java Server Faces – Horstmann, C. S.; Geary, D. M. Título: Core Java Server Faces. Edição: 2ª. Local de publicação: **Alta Books**, 2007. Nº de páginas: 544.
- [4] Icesoft Technologies Inc. IceFaces. Disponível em: <<http://www.icefaces.org>>. Data de acesso: 23/04/2010.
- [5] Entrez Programming Utilities – NCBI. Eutils. Disponível em: <<http://eutils.ncbi.nlm.nih.gov>>. Data de acesso: 23/04/2010.
- [6] eXtensible Markup Language – Ray, Erik T. Título: Learning XML, Second Edition. Edição: 2ª. Local de publicação: **Oreilly & Associates Inc**, 2003. Nº de páginas: 432.