

Desenvolvimento de um programa em Python para extração automatizada de dados na Biblioteca Cochrane: um estudo descritivo

EIXO 2: IMPLEMENTAÇÃO DE TECNOLOGIAS E DIRETRIZES CLÍNICAS EM SAÚDE

Autores: Eduardo Loeffel Noce Gobi; Paola Zucchi; Rafael Leite Pacheco

Introdução: A Biblioteca Cochrane é uma referência global no apoio à tomada de decisões em saúde englobando ensaios e respostas clínicas. A síntese de vários estudos científicos de qualidade, selecionados por critérios de inclusão e exclusão, torna possível a elaboração de uma revisão sistemática. Atualmente, os estudos publicados no Cochrane Database of Systematic Reviews (CDSR) são disponibilizados virtualmente através da internet, sendo o meio pelo qual pesquisadores buscam, avaliam e selecionam esses estudos, para a elaboração de uma revisão sistemática. Entretanto, um dos fatores que leva à demora na seleção de estudos é ter que abri-los um por um dentro das palavras-chave, o que potencializa uma seleção quantitativamente menor. O presente estudo propôs o desenvolvimento de um programa de computador capaz de navegar por todo o CDSR e exportar parágrafos e todos os outros elementos dos estudos em um banco de dados estruturado e direcionado ao pesquisador.

Métodos: O estudo foi conduzido no Núcleo de Avaliação de Tecnologias em Saúde (NATS) da Associação Paulista para o Desenvolvimento da Medicina (SPDM) e segmentado em duas partes, sendo a primeira o protótipo do programa de computador através de scripts nos navegadores Web e a última a consolidação do programa final, escrito na linguagem de programação Python e compatível com os principais sistemas operacionais do mercado.

Resultados: Como teste final, o programa de computador desenvolvido foi capaz de extrair todos os nove mil artigos presentes no CDSR e exportá-los num banco de dados preparado para ser visualizado com programas de planilhas, como o Microsoft Excel. Com o sucesso, o núcleo aprovou o programa.

Discussão e conclusões: O programa passou a ser utilizado como auxílio na seleção de estudos pelos pesquisadores do núcleo, os quais puderam solicitar as extrações com base em palavras-chave ou termos dispostos em quaisquer parágrafos dos textos dos estudos. A extração automática de todos os estudos presentes no CDSR levou cerca de um dia, sem requerer qualquer intervenção dos usuários, um tempo anteriormente tido como “impensável” pelos pesquisadores encarregados de fazer a busca manual dos estudos.

Palavras-chave: Python; Evidence Synthesis; Data Extraction