RESUMO - PRÁTICAS INOVADORAS - INOVAÇÃO E TECNOLOGIA

MABEL: UMA FERRAMENTA DE EXTRAÇÃO DE INFORMAÇÃO PARA AUXÍLIO À REVISÃO SISTEMÁTICA

Gabriel Liston De Menek (gabriellistondemenek@gmail.com)

Gratchela Dutra Rodrigues (gratirodrigues.gdr@gmail.com)

Darling De Andrade Lourenço (darlinglourenco@gmail.com)

Isadora Leitzke Guidotti (leitzke.gi@gmail.com)

Frederico Schmitt Kremer (fred.s.kremer@gmail.com)

A prospecção de novos fármacos nos permite a identificação e caracterização de moléculas com potencial de modular enfermidades. Entre as opções disponíveis atualmente, a busca sistemática por alvos moleculares na literatura científica é uma forma interessante de identificar esses alvos (1). A revisão sistemática de artigos envolve a aplicação de métodos reprodutíveis para encontrar, selecionar e sintetizar todas as evidências disponíveis, de acordo com critérios pré-estabelecidos (2), facilitando o acesso eficiente às evidências de múltiplos estudos. No entanto, o desenvolvimento dessas revisões é um processo demorado, podendo levar de 6 a 18 meses para ser finalizado (3). Idealizado no contexto das demandas na área de prospecção de fármacos, o objetivo deste trabalho é desenvolver o Mapping Biological Entities from Literature (MaBEL), plataforma para a triagem de artigos com a capacidade de extrair automaticamente informações relevantes, como genes e alvos moleculares. Essa abordagem oferece um método mais eficiente para o desenvolvimento de revisões sistemáticas, reduzindo o tempo de realização

das etapas iniciais e fornecendo as informações já sumarizadas. A aplicação foi desenvolvida em linguagem de programação Python, utilizando diferentes bibliotecas para extrair artigos dos principais repositórios de literatura científica, incluindo PubMed, Elsevier, SciELO e arquivos de preprints. Após a busca, informações como título, resumo, autores, páginas, revista onde foi publicado, tipo de artigo, Digital Object Identifier (DOI), afiliações, entre outras, são organizadas em um dataframe padronizado. Isso permite visualizar os dados de artigos provenientes de diferentes fontes em um único arquivo, de forma organizada e unificada. Além disso, para enriquecer o conteúdo apresentado, outras bibliotecas foram implementadas para analisar os dados obtidos e complementá-los com informações e métricas relevantes ao pesquisador. Entre essas bibliotecas, destaca-se o SpaCy, que é capaz de identificar termos relacionados à biologia, como genes e produtos de genes. Além disso, a ocorrência dos termos identificados é analisada e representada em graficos produzidos com a biblioteca Plotly, permitindo destacar alvos moleculares de maior relevância. Para o desenvolvimento da interface front-end da aplicação, foi utilizado o framework Flask. Atualmente, o MaBEL possui toda a estrutura básica para a busca de artigos e visualização dos resultados em dataframes e gráficos de métricas, com todas as funcionalidades da ferramenta disponíveis em uma interface web intuitiva. A busca permite ao usuário construir queries para consulta com palavras-chave e filtros, além de selecionar os repositórios de onde deseja extrair os dados, o número de artigos e que tipo de informações a serem analisadas por meio das funcionalidades de reconhecimento de entidades (NER). Após a requisição da busca os resultados podem ser acessados no histórico do usuário, com opções para visualização online ou download em arquivos em formato de comma separated values (CSV). Dessa forma, o MaBEL é uma plataforma com potencial para acelerar o desenvolvimento de análises sistemáticas, sendo especialmente útil ao levantamento de alvos relevantes, principalmente na prospecção de fármacos. A plataforma reúne todas as funções essenciais para a extração de artigos e a organização dos resultados de forma automatizada, economizando tempo e recursos para o pesquisador e trazendo implicações práticas diretas na otimização dos processos de pesquisa.

Palavras-chave: revisão sistemática; pesquisa científica e desenvolvimento tecnológico; descoberta de drogas.