

## MABEL: UMA FERRAMENTA DE EXTRAÇÃO DE INFORMAÇÃO DA LITERATURA CIENTÍFICA E SEUS AVANÇOS

GABRIEL LISTON DE MENEK<sup>1</sup>; GRACHELA DUTRA RODRIGUES<sup>2</sup>; DARLING DE ANDRADE LOURENÇO<sup>3</sup>; FREDERICO SCHMITT KREMER<sup>4</sup>

<sup>1</sup> Omixlab, Universidade Federal de Pelotas – gabriellistondemenek@gmail.com

<sup>2</sup> Omixlab, Universidade Federal de Pelotas – gratirodrigues.gdr@gmail.com,

<sup>3</sup> BIOSCIENT; Omixlab, Universidade Federal de Pelotas – darlinglourenco@gmail.com

<sup>4</sup> Omixlab, Universidade Federal de Pelotas – fred.s.kremer@gmail.com

### 1. DESCRIÇÃO DA INOVAÇÃO

*Mapping of Biological Entities from Literature* (MaBEL), idealizado para auxiliar no desenvolvimento de revisões sistemáticas, é uma plataforma inovadora para automatizar o processo de triagem de artigos científicos, incluindo a identificação de genes, com um design simples e intuitivo em interface web. A aplicação faz uso do framework Flask e diversas bibliotecas capazes de se conectar, através de Interfaces de programação de aplicações (APIs), aos principais bancos de dados de literatura científica. Dentre as bases estão: Pubmed, Scopus, Science Direct, Scielo e repositórios de *preprints*.

Com as informações extraídas através do MaBEL conseguimos obter e processar grandes quantidades de artigos (até 5000 por repositório), estruturando em um objeto *Dataframe* da biblioteca Pandas. A plataforma oferece visualização organizada dos dados em uma interface moderna, atraente e responsiva. Além disso, conta com pós-processamento por reconhecimento de entidades nomeadas (NER), uma técnica essencial do campo da bioinformática, que permite identificar automaticamente genes e termos biológicos relevantes, e gera gráficos automáticos da ocorrência desses termos, facilitando a interpretação dos resultados (LEE; OH; JUNG, 2023). A integração com o Gemini permite que os dados dos artigos sejam utilizados por um modelo de linguagem, permitindo ao usuário fazer perguntas e obter respostas baseadas nas informações extraídas (TEAM, G. *et al.*, 2023). Essas funcionalidades posicionam a plataforma como uma solução inovadora, destacando-se no mercado pela sua eficiência e avançados recursos de análise.

Title	DOI	Abstract	Date	Human genes	biotator_gene	GENE_OR_GENE_PRODUC
The emerging role of T helper 9 (Th9) cells in immunopathophysiology: A comprehensive review of their effects and responsiveness in various disease states.	<a href="https://doi.org/10.1080/08830185.2024.2364586">10.1080/08830185.2024.2364586</a>	Th9 cells, a subset of T-helper cells producing interleukin-9 (IL-9), play a vital role in the adaptive immune response and have diverse effects in different diseases. Regulated by transcription factors like PU.1 and IRF4, and cytokines such as IL-4 and TGF-β, Th9 cells drive tissue inflammation. This review focuses on their emerging role in immunopathophysiology. Th9 cells exhibit immune-mediated cancer cell destruction, showing promise in glioma and cervical cancer treatment. However, their role in breast and lung cancer is intricate, requiring a deeper understanding of pro- and anti-tumor aspects. Th9 cells, along with IL-9, foster T cell and immune cell proliferation, contributing to autoimmune disorders. They are implicated in psoriasis, atopic dermatitis, and infections. In allergic reactions and asthma, Th9 cells fuel pro-inflammatory responses. Targeting Foxo1 may regulate innate and adaptive immune responses, alleviating disease symptoms.	2024-06-12 00:00:00	IL4, SPI1, IRF4, IL9	IRF4, TGFB1, FOXO1, SPI1, IL4, IL9	IL-9, IRF4, pro-, IL-4, interleukin, PU.1
Advances in CAR-NK cell therapy for lung cancer: is it a better choice in the future?	<a href="https://doi.org/10.3389/fonc.2024.1390006">10.3389/fonc.2024.1390006</a>	Lung cancer remains one of the leading causes of cancer-related mortality worldwide necessitating the development of innovative therapeutic strategies. Chimeric antigen receptor (CAR) natural killer (NK) cell therapy represents a promising advancement in the field of oncology offering a novel approach to target and eliminate tumor cells with high specificity and reduced risk of immune-related adverse effects. This paper reviews the mechanism, potential targets, and recent advances in CAR-NK cell therapy for lung cancer, including the design and engineering of CAR-NK cells, preclinical studies, and the outcomes of early-phase clinical trials. We highlight the unique advantages of using NK cells, such as their innate ability to recognize and kill cancer cells and their reduced potential for inducing graft-versus-host disease (GVHD) and cytokine release syndrome (CRS) compared to CAR T-cell therapies. Results from recent studies demonstrate significant outcomes achieved in lung cancer models with improved targeting and	2024-01-01 00:00:00	CXADR1, TBXT	NR113	Chimeric antigen receptor, C
Effects of metformin on cancers in experimental and clinical studies: Evidence on	<a href="https://doi.org/10.1002/cbf.4071">10.1002/cbf.4071</a>	Metformin (MET) is a preferred drug for the treatment of type 2 diabetes mellitus. Recent studies show that apart from its blood glucose-lowering effects, it also inhibits the development of function tumours, including autochthonous	2024-06-01 00:00:00	PRKAB1, NUP62, PIK3CD, SQSTM1, DNMT	BECN1, PIK3CD, PRKAA2, CCM2, MTO2	p62/SQSTM1, AMPK, adeno: monophosphate-activated p Radin-1, cGMP, DNMT

Figura 1. Resultado de sumarização de artigos na plataforma MaBEL.

## 2. ANÁLISE DE MERCADO

O público alvo do MaBEL inclui empresas de biotecnologia, pesquisadores acadêmicos e laboratórios que frequentemente buscam por inovações e soluções baseadas em evidências da literatura científica, ou buscam otimizar o processo de revisão sistemática. As necessidades específicas atendidas pela plataforma incluem a extração eficiente de dados relevantes, a economia de tempo nas etapas iniciais de pesquisa e a facilitação do acesso a informações cruciais sobre genes e alvos moleculares.

No que tange à concorrência, previamente, outras ferramentas semelhantes já foram publicadas. No entanto, cada uma delas possui suas limitações, algumas por possuírem um sistema de busca exclusivo a apenas um dos repositórios de literatura científica disponíveis, como o *Litstudy* e o *Litreviewer*, outras por não possuírem a capacidade de extração automatizada de artigos, como o *ASReview* e o *Bibliometrix*.

Dados de 2019, considerando os Estados Unidos, mostram que a execução de revisões sistemáticas possuem um custo significativo para as instituições de pesquisa e indústrias farmacêuticas, de US\$18,660,304.77 e US\$16,761,234.71, respectivamente. Além disso, essa forma de revisão pode levar até 18 meses para ser concluída, com aproximadamente 60% do tempo sendo dedicado à busca e avaliação de estudos a serem incluídos (MICHELSON; REUTER, 2019). Nesse contexto, o MaBEL se apresenta como uma ferramenta para auxiliar os pesquisadores, uma vez que torna a busca mais objetiva. Ademais, o formato de comercialização por assinatura reduz os custos associados à compra ou desenvolvimento de soluções pelos pesquisadores/empresas interessados.

O MaBEL possui um potencial de mercado significativo, visto que, o mercado global de biotecnologia teve um tamanho global estimado em US\$ 1.023,92 bilhões, no ano de 2022, com uma previsão de crescimento para US\$ 10.129,16 bilhões em 2030 (“Biotechnology Market Size, Research, Statistics & Industry Outlook By 2030”, 2022). Dentro desse contexto, o mercado brasileiro de biotecnologia possui uma estimativa de US\$ 27.347,3 milhões (“Brazil Biotechnology Market Size & Outlook, 2030”, [s.d.]), com um mercado acessível focado nas empresas que buscam inovações na sumarização da literatura científica e podem se interessar na adoção da plataforma em seus trabalhos.

## 3. ESTRATÉGIA DE DESENVOLVIMENTO E IMPLEMENTAÇÃO

De acordo com o caráter inovador da aplicação desenvolvida, o pedido de registro de software foi submetido e aprovado em setembro de 2023. A aplicação encontra-se em seu estágio final de desenvolvimento, com pequenos ajustes finais planejados antes que seja disponibilizada online para seus primeiros testes de mercado. Sendo assim, o produto se encontra em nível de maturidade tecnológica 4, o que significa que a aplicação já passou pelas fases iniciais de concepção e desenvolvimento, tendo sido validada em ambiente de laboratório.

No entanto, o principal desafio atual do MaBEL reside na relevância de seus resultados. Devido à natureza do método utilizado, que se baseia no reconhecimento de termos biológicos e na contagem de suas ocorrências, é comum que termos genéricos apareçam no topo das listas para a maioria das buscas. Por isso, estamos explorando métodos para aprimorar a qualidade desses resultados e torná-los mais específicos e relevantes.

O MaBEL será comercializado no modelo *Software as a Service* (SaaS) através de assinaturas (mensal, trimestral, semestral e anual) para qualquer indivíduo interessado. Pesquisadores membros de instituições de ensino e pesquisa brasileiras terão descontos de até 30% nas assinaturas e indivíduos poderão fazer até 5 consultas gratuitas por mês. As consultas gratuitas também visam converter o usuário a cliente. Além disso, emails institucionais e publicações em redes sociais serão utilizados como estratégias de marketing. A fidelização dos usuários se dará através de benefícios, como acesso a treinamentos e descontos exclusivos.

#### 4. RESULTADOS ESPERADOS E IMPACTO

Espera-se que o MaBEL impulse o processo de revisão sistemática e análise da literatura científica, alcançando um crescimento gradual na base de usuários. Considerando que nossos concorrentes comercializam suas soluções em valores variando de R\$110 a R\$330 e que nossa meta é obter e manter, ao menos, 10 assinaturas mensais, a ferramenta passará a gerar lucro entre 6 a 16 meses após seu lançamento. Sendo assim, as projeções indicam um retorno positivo já nos primeiros anos de operação. Entretanto, nenhum concorrente possui uma solução similar ao MaBEL, desta forma, estudos de precificação mais aprofundados podem ser necessários.

Futuramente, a aplicação ainda poderá evoluir com novas atualizações, principalmente por meio de implementações que auxiliem na sumarização dos resultados, seja por novos métodos de interpretação dos metadados extraídos, ou por mais funcionalidades de geração de gráficos que auxiliam na visualização dos resultados obtidos.

#### 5. CONCLUSÕES

O MaBEL representa uma ferramenta eficiente para automatizar e otimizar a triagem de artigos científicos. Sua interface intuitiva, com acesso a diversos bancos de dados da literatura e aliada a recursos avançados como NER e geração automática de gráficos, torna o processo de revisão, extração e análise de dados mais ágil e preciso, beneficiando pesquisadores e empresas de biotecnologia na exploração de literatura científica.

Com o produto em fase final de desenvolvimento e um registro de software já aprovado, o MaBEL está pronto para ser testado e aperfeiçoado. Sendo assim convidamos investidores, parceiros e stakeholders a se juntarem a esta iniciativa, contribuindo para o avanço da plataforma e, assim, potencializando a inovação no campo da revisão sistemática e exploração da literatura científica.

#### 6. REFERÊNCIAS BIBLIOGRÁFICAS

- GUIDOTTI, I. L. *et al.* Bambu and its applications in the discovery of active molecules against melanoma. **Journal of Molecular Graphics and Modelling**, v. 124, p. 108564, 1 nov. 2023.
- GOPALAKRISHNAN, S.; GANESHKUMAR, P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. **Journal of Family Medicine and Primary Care**, v. 2, n. 1, p. 9–14, jan. 2013.
- KOUTROUMPA, N.-M. *et al.* A Systematic Review of Deep Learning Methodologies Used in the Drug Discovery Process with Emphasis on In Vivo

Validation. **International Journal of Molecular Sciences**, v. 24, n. 7, p. 6573, jan. 2023.

O'DWYER, L. C.; WAFFORD, Q. E. Addressing challenges with systematic review teams through effective communication: a case report. **Journal of the Medical Library Association : JMLA**, v. 109, n. 4, p. 643–647, 2021.

LEE, S. A.; OH, S.; JUNG, W. **Enhancing Low-resource Fine-grained Named Entity Recognition by Leveraging Coarse-grained Datasets**. arXiv, , 13 nov. 2023. Disponível em: <<http://arxiv.org/abs/2310.11715>>. Acesso em: 7 out. 2024

MICHELSON, M.; REUTER, K. The significant cost of systematic reviews and meta-analyses: A call for greater involvement of machine learning to assess the promise of clinical trials. **Contemporary Clinical Trials Communications**, v. 16, p. 100443, 25 ago. 2019.

TEAM, G. et al. **Gemini: A Family of Highly Capable Multimodal Models**. Disponível em: <<https://arxiv.org/abs/2312.11805v4>>. Acesso em: 7 out. 2024.

**Biotechnology Market Size, Research, Statistics & Industry Outlook By 2030**. Disponível em: <<https://www.databridgemarketresearch.com/reports/global-biotechnology-market>>. Acesso em: 7 out. 2024.

**Brazil Biotechnology Market Size & Outlook, 2030**. Disponível em: <<https://www.grandviewresearch.com/horizon/outlook/biotechnology-market/brazil>>. Acesso em: 7 out. 2024.