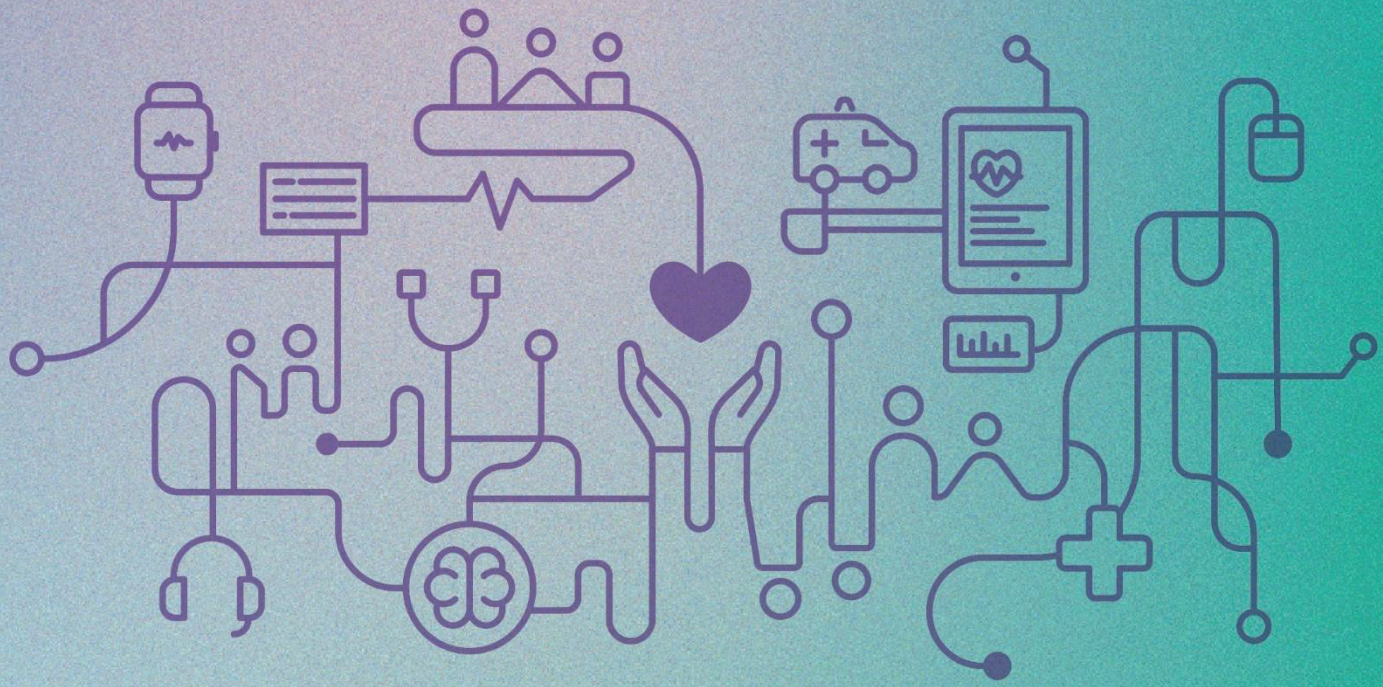
CBIS'24

XX Congresso Brasileiro de Informática em Saúde

SBIS Sociedade Brasileira de Informática em Saúde

ANAIS ESTENDIDOS DO XX CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

SAÚDE DIGITAL: SAÚDE PARA TODOS



08 a 11 de outubro de 2024

Belo Horizonte / MG • Brasil

Sociedade Brasileira de Informática em Saúde.

Anais do XX Congresso Brasileiro de Informática em Saúde : 08 a 11 de outubro de 2024, Belo Horizonte, MG, Brasil / Coordenação da Comissão Científica Juliana Pereira de Souza-Zinader... [*et al.*] . – Sociedade Brasileira de Informática em Saúde (SBIS). – São Paulo, 2024.

501 p. : il.

ISSN: 2178-2857

1. Informática em Saúde 2. Congresso Brasileiro de Informática em Saúde I. Souza-Zinader, Juliana Pereira de. II. Ribeiro-Rotta, Rejane Faria. III. Miranda, Leandro Costa. IV. Gaspar, Juliano de Souza.

ISSN: 2178-2857
DOI: 10.5281/zenodo.14337625

RPA020 - Mapeamento de termos SIGTAP para vocabulários *Observational Medical Outcomes Partnership* (OMOP) por agentes baseados em *large language models*

Vinícius João de Barros Vanzin[1], Dilvan de Abreu Moreira[1]

[1]*Universidade de São Paulo, São Carlos (SP), Brasil.*

vinicius.vanzin@usp.br, dilvan@icmc.usp.br

Link: github.com/vvanzin/sigtap-omop-mapping

Resumo. *Para viabilizar estudos observacionais de larga escala, é necessário que dados de saúde estejam padronizados. Este estudo avalia o mapeamento entre os termos da Tabela do Sistema de Gerenciamento de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais (SIGTAP), utilizado pelo Sistema Único de Saúde (SUS) no Brasil, e os conceitos dos vocabulários do padrão internacional Observational Medical Outcomes Partnership (OMOP), mantido pela Observational Health Data Sciences and Informatics (OHDSI), através de agentes baseados em grandes modelos de linguagem (LLMs na sigla em inglês) com ferramentas para interação com sistemas externos (interfaces Web e bancos vetoriais). Os mapeamentos identificados foram comparados a um conjunto de referência elaborado por especialistas, focando em dois subconjuntos do vocabulário-origem (Procedimentos e Medicamentos) e considerando três métricas: correspondências exatas, parciais e de hierarquia. Na melhor configuração, os agentes atingiram 70% e 80% de acurácia combinada para Procedimentos e Medicamentos, respectivamente. Esses resultados indicam que abordagens baseadas em grandes modelos de linguagem têm o potencial de reduzir o esforço manual dos especialistas no processo de mapeamento.*

Abstract. *To enable large-scale observational studies, healthcare data must be standardized. This study evaluates the mapping between the terms of the Table of the Management System of Procedures, Medications, Orthoses, Prostheses and Special Materials (SIGTAP), used by the Brazilian Unified Health System (SUS), and the concepts of the international Observational Medical Outcomes Partnership (OMOP) standard, maintained by the Observational Health Data Sciences and Informatics (OHDSI), through large language model (LLM) based agents equipped with tools for interacting with external systems (Web interfaces and vector stores). The identified mappings were compared against a reference set developed by specialists, focusing on two subsets of the source vocabulary (Procedures and Medications), and considering three metrics: exact, partial, and hierarchical matches. In the best configuration, the agents achieved 70% and 80% combined accuracy for Procedures and Medications, respectively. These results indicate that LLM-based approaches have the potential to reduce the manual effort of specialists in the mapping process.*

Palavras-chave: Mapeamento de Vocabulários; Large Language Models; Agentes.

1. CONTEXTO E MOTIVAÇÃO

A realização de pesquisas observacionais em larga escala requer que dados clínicos de fontes diferentes estejam representados de maneira consistente e padronizada (1).

Internacionalmente, o modelo OMOP, mantido pela iniciativa OHDSI, busca padronizar o formato dos dados clínicos e seu conteúdo através de vocabulários controlados, permitindo pesquisas colaborativas de larga escala (1). No Brasil, a Tabela do SIGTAP abrange informações dos procedimentos e

DOI: 10.5281/zenodo.14337625

medicamentos utilizados pelo SUS, para fins de faturamento hospitalar e auditorias (2).

O mapeamento dos termos do SIGTAP com os conceitos dos vocabulários OMOP permite que os dados de saúde do Brasil possam ser processados pelas ferramentas do OMOP e utilizados para estudos observacionais (2).

Tradicionalmente, o mapeamento de conceitos entre diferentes recursos terminológicos tem sido abordado por métodos léxicos (análise de elementos textuais dos conceitos), estruturais (hierarquia e relacionamentos entre conceitos) e semânticos (inferência do significado dos conceitos) (3). Devido ao desempenho notável em processamento de linguagem natural, os LLMs, modelos generativos treinados em quantidades massivas de dados (4), têm sido aplicados com sucesso ao mapeamento de vocabulários. Agrawal *et al*. demonstraram que LLMs gerais superam técnicas especializadas na tarefa de extração de informações de dados clínicos, obtendo escore F1 acima de 0,75 para identificação de medicamentos (5). Hertling e Paulheim propuseram uma abordagem combinando filtragem por similaridade semântica e seleção final pelo LLM, alcançando escore F1 dentre os três melhores métodos nas terminologias avaliadas (6).

Nesse contexto, o objetivo deste trabalho é avaliar o uso de agentes baseados em LLMs para o mapeamento de termos do SIGTAP com conceitos dos vocabulários OMOP.

2. DESCRIÇÃO DA SOLUÇÃO

Nossa proposta aborda o mapeamento sob a perspectiva de identificar correspondências pela similaridade semântica dos termos, por meio de agentes baseados em LLMs.

Conjuntos de dados

O conjunto-origem SIGTAP (agosto de 2023) contém 4.731 termos (4.362 Procedimentos e 369 Medicamentos). No conjunto-alvo OMOP, foram considerados conceitos padrão e válidos dos vocabulários SNOMED para Procedimentos e RxNorm/Extension para Medicamentos.

As correspondências determinadas pelos agentes foram comparadas a um conjunto de referência elaborado por especialistas da Comunidade OHDSI Brasil e Sociedade Brasileira de Informática na Saúde (SBIS), engajando dezenas de voluntários. O processo envolveu familiarização inicial, seguida de mapeamento sistemático de Medicamentos, com sessões de revisão para garantir acurácia e consenso entre os participantes (2).

O conjunto de referência aprovado e revisado inclui 487 Procedimentos para SNOMED e 311 Medicamentos para RxNorm. Utilizou-se a tipologia de equivalência do Health Level 7 (7), considerando a intersecção entre mapeamentos em casos de múltiplos conceitos-alvo.

Estrutura dos agentes

Os agentes baseados em LLM para o mapeamento de vocabulários, conforme a abordagem ReAct, recebem uma observação a cada etapa e escolhem uma ação dentro do seu espaço de ações. Uma ferramenta, nesse espaço, refere-se a um sistema externo acessado via formato de invocação pré-definido no prompt de entrada. A saída gerada pela ferramenta é inserida na próxima entrada como uma observação, repetindo a sequência até que o agente atinja uma resposta final (8). O prompt inclui a descrição do cenário, passos a seguir, critérios de uso das ferramentas, formato de saída e o termo-origem, em uma abordagem chain-of-thought (9), na qual o problema é dividido em subtarefas. Para Procedimentos, inclui-se sua descrição. Termos e descrições foram traduzidos para inglês via Google Translate.

DOI: 10.5281/zenodo.14337625

Tabela 1 – Resultados dos experimentos

LLM	Dados	Ferramenta	Saídas válidas	Corr. exatas	Corr. parciais	Corr. hierarquia	Corr. totais
gemini-pro-1.5	Proc.	API	50	6	8	**10**	24 (48%)
gemini-pro-1.5	Proc.	VS	50	8	**18**	9	**35 (70%)**
gpt-4o	Proc.	API	49	13	6	7	26 (53%)
gpt-4o	Proc.	VS	50	**15**	9	6	30 (60%)
Baseline	Proc.	API	50	5	4	4	13 (26%)
Baseline	Proc.	VS	50	6	9	8	23 (46%)
gemini-pro-1.5	Med.	API	33	**19**	0	7	26 (78,8%)
gemini-pro-1.5	Med.	VS	38	**19**	1	6	26 (68,4%)
gpt-4o	Med.	API	50	15	**14**	7	36 (72%)
gpt-4o	Med.	VS	50	13	**14**	13	**40 (80%)**
Baseline	Med.	API	50	11	2	15	28 (56%)
Baseline	Med.	VS	50	12	0	**19**	31 (62%)

Os agentes têm acesso a duas ferramentas: um endpoint da application programming interface (API) OHDSI Athena para consulta aos vocabulários OMOP ou um banco vetorial (VS na sigla em inglês) com os termos-alvo. Em ambas, o agente envia o termo e recebe os códigos e nomes dos primeiros 15 resultados. A ferramenta VS permite acesso a embeddings textuais dos conceitos OMOP, comparáveis por distância euclidiana, extraídos pelo modelo GTE (10) e armazenados no VS FAISS (11). Na execução, o agente consulta a ferramenta, avalia os resultados, e retorna código, nome e tipo de equivalência. Caso nenhum resultado seja adequado, o agente pode alterar o termo e repetir a consulta, até encontrar uma correspondência. O parâmetro de temperatura dos LLMs foi ajustado para 0, minimizando a aleatoriedade da geração.

Métricas de avaliação

Correspondência exata indica que o mapeamento escolhido pelo agente coincide com o conceito-alvo e tipo de equivalência da referência. Correspondência parcial coincide somente em conceito-alvo. Correspondência de hierarquia representa que a escolha do agente é ancestral ou descendente do conceito de referência, segundo a relação de subsunção entre os conceitos na hierarquia.

Por exemplo, para o termo SIGTAP "Microneurólise de nervo periférico", em que o conceito de referência é "SNOMED 4225902 *Neurolysis of peripheral nerve* (WIDER)":

- **Exata:** 4225902 (WIDER);
- **Parcial:** 4225902 (EQUAL);
- **Hierarquia:** 4223677 *Neurolysis* (WIDER), uma vez que 4223677 é ancestral de 4225902.

Resultados e discussão

Foram conduzidos experimentos com 50 termos selecionados aleatoriamente de cada subconjunto (idênticos para cada configuração experimental), variando-se o LLM (GPT-4o ou Gemini Pro 1.5) e a ferramenta disponível. Os resultados estão descritos na Tabela 1, onde "Saídas válidas" indica as execuções com correspondências no formato esperado. Saídas inválidas ocorreram por excesso de chamadas de ferramenta, quebra de protocolo, ou interrupção devido às políticas de conteúdo.

O desempenho dos agentes foi comparado a *baselines* que consideram como conceito-alvo o primeiro resultado retornado da API ou VS, com o tipo de equivalência mais frequente, isto é, consultas tradicionais sem o uso de agentes baseados em LLMs. Todos os agentes superaram os *baselines* em seus respectivos

DOI: 10.5281/zenodo.14337625

conjuntos. O agente com LLM Gemini e ferramenta VS atingiu 70% de acurácia para Procedimentos, enquanto o agente com LLM GPT-4 e VS obteve 80% para Medicamentos.

Em média, a diferença de desempenho entre as ferramentas é menor para Medicamentos, indicando que consultas léxicas (API) são suficientes para medicamentos, enquanto procedimentos requerem análise em nível semântico. A ferramenta API utiliza recursos pré-existentes, enquanto a ferramenta VS requer extração e armazenamento dos *embeddings*.

3. PONTOS RELEVANTES DA INOVAÇÃO

Em comparação a abordagens anteriores que limitam termos-candidatos (5-6), nossa proposta tem acesso efetivamente a todo o vocabulário-alvo durante o mapeamento, uma vez que o agente pode avaliar os resultados das ferramentas e efetuar novas consultas. O agente é capaz de ajustar dinamicamente as consultas futuras até que um mapeamento adequado seja encontrado, isto é, não se limita somente ao termo original para comparação.

A automatização do mapeamento, mesmo que parcial, permite que os especialistas se concentrem em termos que necessariamente exigem maior análise e consenso. Ao aplicar modelos de propósito geral, minimiza-se a necessidade de ajuste fino, podendo ser empregada para outros recursos terminológicos.

Aspectos Éticos

As bases SIGTAP e OMOP estão disponíveis publicamente. Não houve coleta de dados sensíveis ou material biológico. Portanto, não existem aspectos éticos ou conflitos de interesse a serem considerados.

4. CONSIDERAÇÕES FINAIS

Este artigo apresentou uma abordagem de agentes baseados em LLMs para o mapeamento das terminologias SIGTAP e OMOP, com o propósito de conectar o vocabulário brasileiro ao cenário internacional. Na melhor configuração, atingiram-se 70% e 80% de acurácia para Procedimentos e Medicamentos, respectivamente, considerando-se as métricas combinadas. Em trabalhos futuros, planeja-se aumentar o número de termos mapeados, expandir a seleção de ferramentas e explorar LLMs abertos.

AGRADECIMENTOS

Agradecemos ao Prof. Ricardo Marcacini por suas contribuições e à empresa Sofya pelos créditos para uso do LLM GPT-4. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

1. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, *et al*. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *In:* Studies in health technology and informatics; 2015. p. 574-8.

2. Gualdani FA, Sanovo O, Abrahão MT, Oliveira NB, Fedel A, Moreira DA, *et al*. Mapeamento de termos SIGTAP para o padrão OMOP: relato de experiência de um grupo de estudos. *In:* Congresso Brasileiro de Informática em Saúde; 2022. p. 155-60.

3. Euzenat J, Shvaiko P. Ontology matching. 2nd ed. Berl*In:* Springer, 2013.

4. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, *et al*. Language models are few-shot learners. Advances in Neural Information Processing Systems. 2020;33:1877-901.

5. Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. *In:* Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; 2022. p. 1998-2022.

6. Hertling S, Paulheim H. Olala: Ontology matching with large language models. *In:* Proceedings of the 12th Knowledge Capture Conference; 2023. p. 131-9.

7. HL7. ConceptMapEquivalence [Internet]. 2024. [citado 25 mai 2024]. Disponível em: https://www.hl7.org/FHIR/DSTU1/concept-equivalence.html.

8. Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan, K, *et al*. ReAct: Synergizing Reasoning and Acting in Language Models. *In:* International Conference on Learning Representations (ICLR); 2023. 33 p.

9. Wei J, Wang X, Schuurmans D, Bosma M, Xia F, Chi E, *et al*. Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems. 2022;35:24824-37.

10. Li Z, Zhang X, Zhang Y, Long D, Xie P, Zhang M. Towards general text embeddings with multi-stage contrastive learning [Internet]. arXiv [Preprint]. 2023 [citado 25 mai 2024]: 18 p. Disponível em: https://arxiv.org/abs/2308.03281.

11. Douze M, Guzhva A, Deng C, Johnson J, Szilvasy G, Mazaré P-E, *et al*. The FAISS library [Internet]. arXiv [Preprint]. 2024: [citado 25 mai 2024]: 21 p. Disponível em: https://arxiv.org/abs/2401.08281.

DOI: 10.5281/zenodo.14337625