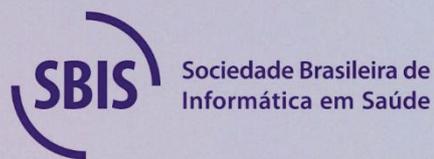


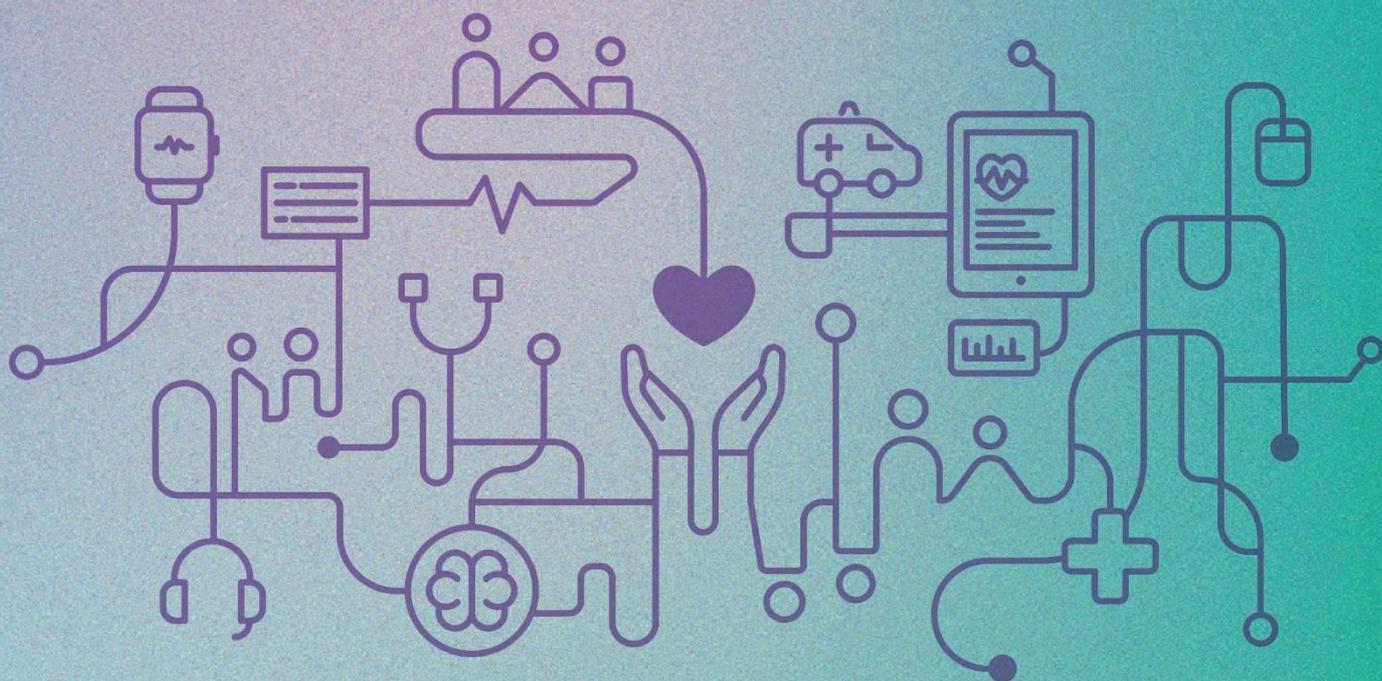
XX Congresso Brasileiro
de Informática em Saúde

CBIS'24



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.

Copyright© 2024 Autores individuais dos textos publicados.

Permitida a redistribuição, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com crédito atribuído aos(às) autores(as).

ISSN: 2178-2857

DOI: 10.5281/zenodo.14337625



RPA011 - Desenvolvimento e utilização de sub-ontologias da SNOMED CT para a identificação de anafilaxia em prontuários

Joice Basilio Machado Marques¹, Reinaldo Porte Peres¹, Fabrício Amandeu Gualdani²,
Matheus Matos Machado³, Monica Pugliese Heleodoro dos Santos¹, Fabio Cerqueira Lario⁴,
Chayanne Andrade de Araujo⁴, Fabiana Aandrade Nunes Oliveira⁴, Luis Felipe Chiaverini Ensina⁴,
Dilvan de Abreu Moreira³

¹Sofya Tecnologia, São Paulo (SP), Brasil.

²Universidade Estadual Paulista UNESP, São Paulo (SP), Brasil.

³Universidade de São Paulo USP, São Paulo (SP), Brasil.

⁴Hospital Sírio-Libanês, São Paulo (SP), Brasil

joicemachado@1sti.com.br; reinaldoperes@1sti.com.br; fabricio.gualdani@unesp.br;
matheusmatos@usp.br; dramonicapugliese@gmail.com; fabio.clario@hsl.org.br;
chayanneaa79@gmail.com; fabiana.andrade.nunes@gmail.com; 100alergia@gmail.com;
dilvan@icmc.usp.br

Resumo. A anafilaxia é uma reação grave que, após a exposição a um alérgeno ou substância, ocasiona rápida progressão, de minutos a poucas horas, podendo levar o paciente a óbito. Este trabalho tem como objetivo apresentar o desenvolvimento de sub-ontologias para auxiliar na identificação e sinalização de ocorrências (presente ou anterior) de uma reação anafilática. A sub-ontologia foi desenvolvida a partir de um conjunto de termos clínicos definidos na terminologia SNOMED CT e validados por especialistas médicos. Além dos termos clínicos, os critérios para identificar o diagnóstico foram estabelecidos, em conformidade com a literatura, para inferir a ocorrência ou não da reação anafilática. A partir da inserção dos dados, a aplicação indica para o profissional se há ou não suspeita de anafilaxia, de acordo com a base de conhecimento estabelecida. Ademais, um formulário será desenvolvido considerando as especificidades do contexto de anafilaxia e o recurso AllergyIntolerance (FHIR).

Abstract. Anaphylaxis is a serious reaction that, after exposure to an allergen or substance, causes rapid progression, lasting minutes to a few hours, and can lead to the patient's death. This work aims to present the development of sub-ontologies to help identify and signal occurrences (present or previous) of an anaphylactic reaction. The sub-ontology was developed from a set of clinical terms defined in SNOMED CT terminology and validated by medical experts. In addition to the clinical terms, the criteria to identify the diagnosis were established, in accordance with the literature, to infer the occurrence or not of the anaphylactic reaction. After entering the data, the application indicates to the professional whether or not anaphylaxis is suspected, according to the established knowledge base. Furthermore, a form will be developed considering the specificities of the anaphylaxis context and the AllergyIntolerance (FHIR) resource.

Palavras-chave: Anafilaxia; Informática clínica; Terminologias; SNOMED CT.



1. CONTEXTO E MOTIVAÇÃO

Em sistemas de apoio à decisão clínica, as ontologias são usadas para expressar estruturas e objetos com interconexões complexas, além de manter flexibilidade e compatibilidade com outros sistemas. Uma ontologia é uma especificação explícita de uma conceituação de um determinado contexto, bem como seus relacionamentos e propriedades. Em sistemas computacionais, a ontologia descreve o conhecimento de um domínio representado em um formalismo declarativo (1).

A SNOMED CT (*Systematized Nomenclature of Medicine – Clinical Terms*) é uma das ontologias médicas mais abrangentes e amplamente utilizadas no mundo. Ela fornece um vocabulário padronizado que permite uma descrição consistente do conteúdo clínico em registros de saúde. Ela foi projetada para ser usada em diferentes ambientes de assistência médica, incluindo prestação de cuidados, suporte a decisões clínicas, análise e pesquisa (2).

Além disso, o uso de terminologias também possibilita vantagens como: rigor clínico já conferido na ontologia; padronização de vocabulários para a construção da interoperabilidade; criação de conjuntos de referências que possam ser utilizados e atualizados por outros grupos; suporte a ferramentas que automatizam o prontuário eletrônico, e utilização de recursos HL7 FHIR (*Health Level Seven International -Fast Health Interoperability Resources*) (3).

O caso clínico abordado está no escopo de reações anafiláticas. A anafilaxia é a apresentação clínica mais grave das reações alérgicas sistêmicas agudas. Os sintomas aparecem após a exposição a um alérgeno ou outro gatilho, com rápida progressão em minutos a poucas horas (4).

Portanto, o objetivo deste projeto é auxiliar na identificação de anafilaxia e evitar a reincidência de casos, melhorando a qualidade de vida do paciente, otimizando atendimento nos hospitais e minimizando os custos de internação.

O presente trabalho está em andamento e a ênfase será dada ao processo de criação das sub-ontologias que serão utilizadas para validar as inferências acerca da suspeita de anafilaxia. As etapas de processamento de linguagem natural e aplicação das técnicas de inteligência artificial são tratadas em outras etapas do projeto de pesquisa em desenvolvimento. Serão considerados os Grandes Modelos de Linguagem (*Large Language Models- LLMs*) e a Geração Aumentada de Recuperação (*Retrieval-Augmented Generation - RAG*) para a extração e validação dos dados clínicos. Na Figura 1 é apresentada, em alto nível de abstração, a proposta da arquitetura utilizada neste projeto e seus principais resultados.

Além do desenvolvimento da sub-ontologia no domínio de anafilaxia, a solução abordada neste artigo tem como interface um formulário de inserção de dados que se baseia no recurso FHIR AllergyIntolerance para a estruturação dos dados. Vale ressaltar que os dados são retrospectivos e foram retirados de uma base pública para fazer os testes na instanciação da ontologia.

Nas próximas seções serão descritas as técnicas utilizadas, resultados preliminares da pesquisa e algumas perspectivas futuras de desenvolvimento.

2. DESCRIÇÃO DA SOLUÇÃO

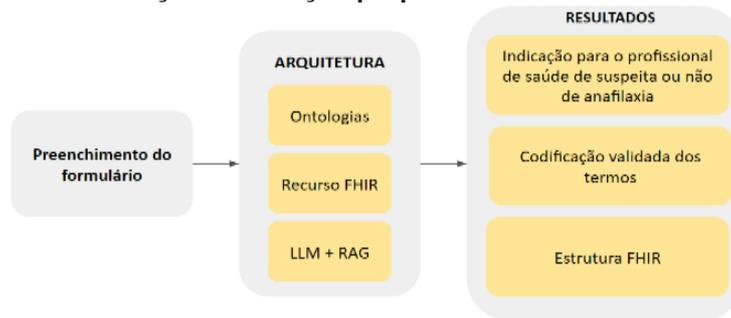
Além das tecnologias mencionadas na seção anterior, este projeto também abrange o uso de modelos de inteligência artificial, que na medicina inclui o uso de algoritmos complexos e software para imitar ou estimar a cognição



humana e a análise de contextos complexos e dados relacionados à saúde (5). Ela tem sido amplamente utilizada nas pesquisas aplicadas à saúde, com o intuito de otimizar a jornada do

cuidado do paciente e de extrair o conhecimento do grande volume de dados com rapidez e eficiência.

Figura 1 - Descrição da solução proposta em alto nível de abstração



Recentemente emergiram os Grandes Modelos de Linguagem, que demonstram capacidades significativas, mas enfrentam desafios como alucinações, conhecimento desatualizado e processos de raciocínio não transparentes. Nesse sentido, a RAG surgiu como uma solução promissora ao incorporar conhecimento de bancos de dados externos (6). Estas bases de conhecimento auxiliam na rastreabilidade e na recuperação de conteúdo validado e codificado por especialistas, como no caso deste projeto.

Aliado aos grandes modelos de linguagem da inteligência artificial, o uso de ontologias médicas pode melhorar significativamente a qualidade dos dados e o impacto da pesquisa biomédica, conforme apresentado na seção anterior. Estas podem ser usadas para representar e analisar uma grande variedade de dados clínicos sobre sintomas, diagnósticos e tratamentos de pacientes, por exemplo (7).

Apesar desses avanços, os LLMs apresentam limitações, particularmente no tratamento de consultas específicas de domínio ou altamente especializadas (6). Por isso, esta abordagem utilizará o RAG que será alimentado, dentre outras bases de conhecimento, pela sub-

ontologia da SNOMED CT para obter as recomendações em conformidade com as entidades mapeadas e codificadas pelos especialistas. Outras técnicas também poderão ser usadas durante o desenvolvimento do trabalho e serão avaliadas no decorrer do projeto.

3. PONTOS RELEVANTES DA INOVAÇÃO

Na recuperação de conteúdo, uma metodologia híbrida que aproveita fontes de dados estruturadas e não estruturadas está emergindo como uma forte tendência, proporcionando um processo de recuperação mais enriquecido (6). Estes tipos de aplicações são muito úteis aos cenários clínicos, onde tanto os formulários estruturados quanto as notas discursivas sobre a evolução do paciente apresentam dados significativos para compor o sumário do paciente, por exemplo.

No caso clínico de anafilaxia, o reconhecimento imediato dos sintomas que potencialmente colocam os pacientes em risco aumentado de anafilaxia grave é obrigatório para o sucesso do tratamento e diminuição da taxa de mortalidade, principalmente no âmbito hospitalar. Além disso, o diagnóstico clínico



adequado e a identificação do agente causal são fundamentais para a prevenção de novos episódios (8).

Dada a complexidade do cenário, este trabalho apresenta diversas fases de desenvolvimento e validação clínica. Uma equipe de médicos alergistas foi responsável tanto pela definição do escopo do projeto, quanto pela criação do conjunto de sintomas, sinais e outros achados relacionados à anafilaxia.

A primeira etapa refere-se à construção do conjunto de referências acurado pelos especialistas do grupo de pesquisa. Após a definição dos termos, os terminologistas fizeram o mapeamento para os conceitos da SNOMED CT. Os médicos especialistas validaram os termos verificando a correspondência mais acertada para o contexto. Todos estes passos corroboram para a inserção dos conceitos corretos na sub-ontologia e, futuramente, no prontuário do paciente.

A segunda etapa tratou do desenvolvimento de uma ferramenta para auxiliar a construção das sub-ontologias. O time de desenvolvimento aprimorou a ferramenta proposta por Del-Pinto *et al.* (9) para fazer a extração de sub-ontologias da terminologia, a partir de um conjunto de conceitos da SNOMED CT, ou seja, o artefato resultante da primeira etapa.

Vale ressaltar que este resultado parcial é fundamental para o uso da SNOMED CT para soluções de domínio específico. O terminologista insere um conjunto de códigos válidos e a ferramenta extrai classes, axiomas, propriedades e principais relacionamentos gerando uma sub-ontologia da SNOMED CT. Este corte facilita a utilização por domínios específicos e com uma maior eficiência computacional, uma vez que não é preciso executar os mais de 360 mil conceitos da ontologia. A sub-ontologia resultante tem

pouco menos de 3000 conceitos, possui todas as propriedades e os axiomas já definidos para reações adversas e anafilaxia, bem como alguns agentes causadores. Essa sub-ontologia será parte da base de conhecimento aliada ao modelo de linguagem.

A terceira etapa, e fase atual de desenvolvimento, consiste na adaptação do formulário do recurso FHIR e SNOMED CT para sua utilização no contexto de anafilaxia. A partir da inserção dos dados relacionados ao contexto do paciente, a aplicação irá sinalizar a existência ou não da suspeita de anafilaxia. Estas recomendações terão como base a sub-ontologia e serão exportadas seguindo o formato do recurso FHIR.

4. CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentada uma abordagem de construção de sub-ontologias de domínio, com base em uma ferramenta de extração de conceitos da SNOMED CT. A sub-ontologia de anafilaxia, construída com um conjunto de referências validado por médicos especialistas, será parte da base de conhecimento acessada pelo LLM para dar maior acurácia e rastreabilidade nas recomendações propostas pela aplicação que será desenvolvida.

Diagnosticar com maior rapidez a ocorrência de anafilaxia nos prontuários, garantir que os dados chave informados em texto livre sejam capturados e encaminhados para seus respectivos campos estruturados nos prontuários eletrônicos são ações futuras que fazem parte do *roadmap* do projeto, a fim de não se perder informações relevantes sobre a jornada do paciente.

Espera-se aprimorar a identificação de possíveis anafilaxias, com consequente melhora do registro estruturado e redução do impacto de eventos adversos à saúde dos pacientes, que até então são prejudicados por



reexposição aos agentes causadores devido à subnotificação dos eventos.

AGRADECIMENTOS

Este projeto é financiado pelo Instituto de Ensino e Pesquisa (IEP) do Hospital Sírio Libanês e pela empresa Sofya que mantém o grupo de pesquisa aplicada em informática clínica.

REFERÊNCIAS

1. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*. 1995;43(5-6):907-28.
2. INTERNATIONAL S. SNOMED CT Terminology Services Guide - Terminology Services Guide - SNOMED Confluence. confluence.hl7.org. 2022.
3. HL7 FHIR - Fast Health Interoperability Resources. <https://hl7.org/fhir/allergyintolerance.html>
4. Cardona V, Ansotegui IJ, Ebisawa M, El-Gamal Y, Fernandez Rivas M, Fineman S, *et al*. World allergy organization anaphylaxis guidance 2020. *World Allergy Organ J*. 2020;13(10):100472.
5. Deshmukh, R., & Rathi, P. (2022). Artificial Intelligence in Medicine. In *The Journal of the Association of Physicians of India* (Vol. 70, Issue 3).
6. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Guo, Q., Wang, M., & Wang, H. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey.
7. Burgun A, Bodenreider O. Accessing and integrating data and knowledge for biomedical research. *Yearb Med Inform*. 2008;91-101.
8. Ensina LF, Félix MMR, Aranda CS. Drug-induced anaphylaxis: clinical scope, management, and prevention. *Current Treatment Options in Allergy*. 2016; 3:243-52.
9. Del-Pinto, W., Schmidt, R. A., Gao, Y., Alghamdi, G., Osornio, A. L., & Roy, S. (2024). International Patient Summary Terminology. *Studies in Health Technology and Informatics*, 310, 63–67.