UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Marília

ALLAN FERREIRA

Modelo de arquitetura para interoperabilidade de dados de saúde utilizando padrão FHIR

Marília
2024

ALLAN FERREIRA

Modelo de arquitetura para interoperabilidade de dados de saúde utilizando padrão FHIR

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação como parte das exigências para a obtenção do título de Mestre em Ciência da Informação pela Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Campus de Marília.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Leonardo Castro Botega

Coorientadora: Profa. Dra. Mariângela Spotti Lopes Fujita

Marília
2024

F383m

Ferreira, Allan
Modelo de arquitetura para interoperabilidade de dados de saúde utilizando padrão FHIR / Allan Ferreira. -- Marília, 2024
99 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília
Orientador: Leonardo Castro Botega
Coorientadora: Mariângela Spotti Lopes Fujita

1. Arquitetura da Informação. 2. FHIR. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília.
Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa

O modelo FHIR-FLOW impacta diretamente na interoperabilidade entre sistemas de saúde, melhorando a comunicação entre diferentes plataformas de registros eletrônicos de saúde (EHR). Ele utiliza tecnologias avançadas como LLM e BERT dentro do padrão FHIR para análise e mapeamento precisos dos dados de saúde. Essa abordagem não apenas facilita diagnósticos mais precisos e tratamentos personalizados, mas também contribui significativamente para a Ciência da Informação, fornecendo uma base sólida para futuras pesquisas e práticas na área de interoperabilidade de dados de saúde.

Potencial impact of this research

The FHIR-FLOW model directly impacts interoperability between healthcare systems by improving communication between different electronic health records (EHR) platforms. It uses advanced technologies like LLM and BERT within the FHIR standard for accurate analysis and mapping of healthcare data. This approach not only facilitates more accurate diagnoses and personalized treatments, but also contributes significantly to Information Science by providing a solid foundation for future research and practice in the area of health data interoperability.

Impacto potencial de esta investigación

El modelo FHIR-FLOW impacta directamente la interoperabilidad entre los sistemas de salud al mejorar la comunicación entre diferentes plataformas de registros médicos electrónicos (EHR). Utiliza tecnologías avanzadas como LLM y BERT dentro del estándar FHIR para un análisis y mapeo precisos de datos de atención médica. Este enfoque no sólo facilita diagnósticos más precisos y tratamientos personalizados, sino que también contribuye significativamente a las Ciencias de la Información al proporcionar una base sólida para futuras investigaciones y prácticas en el área de la interoperabilidad de datos de salud.

Allan Ferreira

Modelo de arquitetura para interoperabilidade de dados de saúde utilizando padrão FHIR

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), como requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Informação, Tecnologia e Conhecimento
Linha de pesquisa: Informação e Tecnologia

Banca Examinadora

Prof. Dr. Leonardo Castro Botega
UNESP – Câmpus de Marília
Orientador

Prof. Dr. Caio Saraiva Coneglian
Centro Universitário Eurípides de Marília
Examinador

Prof. Dr. Joice Basílio Machado Marques
Pesquisador/Sofya
Examinador

Marília, 01 de março de 2024.

AGRADECIMENTOS

À Deus pela oportunidade e condições de realizar este trabalho.

Aos meus pais, Zilda e Lourenço. Nenhuma palavra humana é capaz de expressar em sua completude os meus sentimentos de amor, carinho e gratidão por tudo o que vocês sempre fizeram por mim.

À minha esposa Sara por todo apoio, incentivo e cuidado com nossa família especialmente nos momentos que precisei alocar energia extra para a construção deste trabalho

Ao meu orientador Prof. Dr. Leonardo Castro Botega, pela oportunidade, paciência e orientações, as quais impactaram minha carreira positivamente de um jeito ímpar.

Ao grupo de estudos “HAIS” do Hospital das Clínicas da Faculdade de Medicina de Marília, pela participação na fase inicial deste trabalho.

Aos colegas do PPGCI da Unesp e dos grupos de pesquisa, por todo apoio, companheirismo e troca de conhecimento.

RESUMO

A evolução dos registros eletrônicos de saúde revelou um desafio crítico diretamente ligado à interoperabilidade dos dados, causado pela falta de padrões unificados entre os sistemas de saúde. Este estudo enfrentou tal desafio ao desenvolver um modelo de arquitetura da informação chamado FHIR-FLOW, que utiliza o padrão *Fast Healthcare Interoperability Resources* (FHIR) e incorpora tecnologias avançadas de aprendizado de máquina, especificamente, Modelos de Linguagem Grande (LLM) e o modelo BERT, para o reconhecimento eficaz de entidades e extração de recursos FHIR. A pesquisa fundamentou-se na análise detalhada de dados de alergia, provenientes dos prontuários eletrônicos do Hospital Sírio Libanês, aliada a uma extensa revisão bibliográfica sobre interoperabilidade, destacando-se pela integração da Ciência da Informação (CI) através do uso de uma matriz de metadados de negócios, que foi importante para estruturar e interpretar os dados de saúde de maneira coerente, assegurando uma base sólida para a construção e validação do modelo. Os resultados obtidos com o FHIR-FLOW indicaram melhorias na interoperabilidade que facilitam a comunicação entre sistemas de saúde diversos. A implementação das técnicas de LLM e BERT, adicionadas do *Snowstorm* (um classificador de termos utilizado na camada terminológica), permitiu uma interpretação e processamento precisos das terminologias médicas. Em cenários generalistas, o modelo alcançou taxas de acerto superiores a 66%, enquanto em cenários específicos de alergia, a precisão ultrapassou os 90%. Tais achados evidenciam a eficácia do modelo não apenas na melhoria da interoperabilidade e na comunicação entre sistemas, e este avanço significativo para os cuidados de saúde na era digital enfatiza a contribuição da Ciência da Informação, especialmente no que tange ao desenvolvimento e aplicação de estruturas de dados. O sucesso do FHIR-FLOW ressalta a importância de abordagens inovadoras na superação dos desafios de interoperabilidade, servindo como referência para futuras pesquisas e práticas na área.

Palavras-chave: Arquitetura da informação; FHIR.

ABSTRACT

The evolution of electronic health records has revealed a critical challenge directly linked to data interoperability, caused by the lack of unified standards across healthcare systems. This study addressed this challenge by developing an information architecture model called FHIR-FLOW, which uses the FHIR (Fast Healthcare Interoperability Resources) standard and incorporates advanced machine learning technologies, specifically Large Language Models (LLM) and the BERT model. for effective entity recognition and FHIR feature extraction. The research was based on the detailed analysis of allergy data from the electronic medical records of Hospital Sírio Libanês, combined with an extensive bibliographic review on interoperability, highlighting the integration of Information Science through the use of a business metadata matrix, It was important to structure and interpret health data in a coherent way, ensuring a solid basis for model construction and validation. The results obtained with FHIR-FLOW indicated improvements in interoperability that facilitate communication between different health systems. The implementation of LLM and BERT techniques, added to Snowstorm (a term classifier used in the terminological layer), allowed accurate interpretation and processing of medical terminologies. In general scenarios, the model achieved accuracy rates above 66%, while in allergy-specific scenarios, accuracy exceeded 90%. Such findings highlight the effectiveness of the model not only in improving interoperability and communication between systems, and this significant advance for healthcare in the digital era emphasizes the contribution of Information Science, especially with regard to the development and application of information structures. data. The success of FHIR-FLOW highlights the importance of innovative approaches in overcoming interoperability challenges, serving as a reference for future research and practices in the area.

Keywords: Information Architecture; FHIR.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema do passo a passo da metodologia 20
Figura 2 – Mapa conceitual do ambiente informacional de prontuário eletrônico 24
Figura 3 – Visão de acesso aos recursos na arquitetura FHIR 32
Figura 4 – Diagrama UML parcial do recurso FHIR *Patient* 33
Figura 5 – Mapa conceitual da arquitetura do padrão FHIR 35
Figura 6 – Etapas do modelo BERT 43
Figura 7 – Machine Learning na área da saúde 49
Figura 8 – Representação, recuperação e acesso à informação de dados clínicos 54
Figura 9 – Mapa conceitual dos atributos do *dataset* de alergia do HSL 57
Figura 10 – Representação do recurso FHIR AllergyIntolerance 58
Figura 11 – Fluxo geral de transformações dos dados e etapas do FHIR-FLOW 66
Figura 12 – Etapas da extração sintática e semântica 69
Figura 13 – Camada de identificação FHIR 73
Figura 14 – Camada terminológica 75
Figura 15 – Fluxo geral do modelo 80
Figura 16 – Amostra de validação de texto livre com a aplicação atuando 91
Figura 17 – Tela de validação de dados estruturados e *download* do resultado 91

Gráfico 1 – Análise de presença e ausência de dados 87

LISTA DE TABELAS

Tabela 1 – Dados comparativos entre trabalhos correlatos 30
Tabela 2 – Principais tipos de Machine Learning e sua correlação com a CI 38
Tabela 3 – Exemplo de prompt de chamada para GPT3.5-Turbo 47
Tabela 4 – Amostra original dos dados de alergia do HSL 56
Tabela 5 – Metadados de negócio gerados com interoperabilidade no padrão FHIR 63
Tabela 6 – Exemplo de extração de triplas a partir de fontes de dados 70
Tabela 7 – Exemplo de identificação de recursos FHIR através de triplas 74
Tabela 8 – Exemplo de identificação terminológica em recursos FHIR 76
Tabela 9 – Uso da SNOMED CT em diversas áreas da saúde 78
Tabela 10 – Anamneses geradas para validação de conformidade FHIR 82
Tabela 11 – Receituários gerados para validação de conformidade FHIR 83
Tabela 12 – Dados de alergia gerados para validação de conformidade FHIR 85
Tabela 13 – Amostra de dados 88
Tabela 14 – Amostra de dados validados 89
Tabela 15 – Dados não validados 90

LISTA DE ABREVIATURAS E SIGLAS

AI	Arquitetura da Informação
API	Application Programming Interface
BERT	Bidirecional Encoder Representations from Transformers
BRAPCI	Biblioteca de Dados em Ciência da Informação
CI	Ciência da Informação
EMR	Electronic Medical Records
EHR	Electronic Health Records
ERP	Enterprise Resource Planning
FIHR	Fast Healthcare Interoperability Resources
HL7	Health Level 7
HSL	Hospital Sírio Libanês
HTTP	Hypertext Transfer Protocol
IA	Inteligência Artificial
ICD	International Classification of Diseases
IEEE	Institute of Electrical and Electronics Enginners
JSON	JavaScript Object Notation
LLM	Large Language Model
LOINC	Logical Observation Identifiers Names and Codes
MIMIC	Medical Information Mart for Intensive Care
ML	Machine Learning
MRCM	Machine-Readable Concept Model
NER	Named Entity Recognition
NLP	Natural Processing Language
PEP	Prontuário Eletrônico do Paciente
PLN	Processamento de Linguagem Natural

RCTM Rede Catarinense de Telemedicina

RDA Resource Description Framework

REST Representational Transfer State

SNOMED CT Systematized Nomenclature of Medicine – Clinical Terms

TISS Troca de Informação para Saúde Suplementar

UML Unified Modeling Language

USI Unidade Semi Intensiva

URI Uniform Resource Identifier

VA Veteran's Administration

XML Extensible Markup Language

SUMÁRIO

1	**INTRODUÇÃO**	15
1.1	**Problema de Pesquisa**	16
1.2	**Justificativa**	17
1.3	**Metodologia**	18
1.4	**Estrutura da Pesquisa**	21
2	**PANORAMA DOS DADOS DE SAÚDE**	23
2.1	**Interoperabilidade de dados de saúde**	26
3	**INTEROPERABILIDADE DE DADOS DE SÁUDE: TRABALHOS CORRELATOS**	28
4	**PADRÃO FHIR**	31
4.1	**Arquitetura Geral do FHIR**	31
4.2	**Recursos FHIR**	32
5	**MACHINE LEARNING NA ÁREA DA SAÚDE**	37
5.1	**Potencial e aplicações**	37
5.2	**Processamento de Linguagem Natural**	40
5.3	**Modelos de Aprendizado de Máquina em PLN: BERT**	42
5.4	**Modelos de Linguagem Grande (LLMs)**	45
5.4.1	Modelo GPT3.5-Turbo	46
5.5	**Desafios na Integração de ML e Saúde**	48
6	**DADOS DE ALERGIA DO HOSPITAL SÍRIO LIBANÊS: ANÁLISE COM METADADOS DE NEGÓCIO**	51
6.1	**Importância de dados de alergia**	51
6.2	**Aplicação de metodologia de Arquitetura da Informação**	53
7	**FHIR-FLOW: MODELO DE MAPEAMENTO DE DADOS DE SAÚDE PARA FHIR**	65
7.1	**Camada Sintática e Semântica**	67
7.2	**Camada Identificação FHIR**	72
7.3	**Camada Terminológica**	74
7.3.1	SNOMED CT	77
7.3.2	Servidor de terminologias Snowstorm	79
7.4	**Validação do modelo**	79
7.4.1	Avaliação da conformidade FHIR	81

7.4.2 Integridade e consistência dos dados 87
8 CONSIDERAÇÕES FINAIS 93
8.1 Conclusões 93
8.2 Contribuições 94
REFERÊNCIAS 96

1 INTRODUÇÃO

Com o aumento do desenvolvimento de aplicações, que permitem a transposição de registros físicos para meios eletrônicos, a área de saúde tem se beneficiado com o aspecto de persistência e recuperação de dados através dos *Eletronic Medical Records* (EMR), na influência direta de como os profissionais ministram os cuidados aos pacientes, e no auxílio do pensamento clínico crítico (Tierney, 2013).

Mediante o número de instituições e *softwares* existentes, um dos desafios dos EMR está na padronização da Arquitetura da Informação (AI), visto que um paciente, ao longo da vida, tem seus dados registrados em diferentes instituições, com diferentes bases de dados e estruturas de armazenamento. As principais dificuldades concentram-se no problema da arquitetura e representação da informação para uso computacional, mediante a complexidade do cenário de saúde, e na existência de um grande volume de padrões e arquiteturas existentes, em que cada instituição tem dificuldade de escolha particular no momento de selecionar a arquitetura que lhe trará um melhor custo-benefício (Petry *et al.*, 2008).

Neste sentido, a interoperabilidade dos dados de saúde, definido como a capacidade de dois ou mais sistemas cooperarem apesar das diferenças de linguagem, estruturas ou plataforma de execução (Wegner, 1996), tem como necessidade uma construção de uma arquitetura da informação que proporcione a troca de dados entre os diferentes sistemas, para permitir o aumento da capacidade de organização e recuperação dos dados. Portanto, pode gerar uma série de benefícios para as organizações de saúde, como cuidado mais eficaz ao paciente e a possibilidade de se recuperar informação de diferentes fontes que estão distribuídas e armazenadas em ambientes heterogêneos (Nardon, 2003).

No sentido de auxiliar a arquitetura da informação de modelos interoperáveis na área da saúde, existe uma série de padrões que norteiam a modelagem de metadados de negócio e facilitam a troca de dados entre instituições que adotam os mesmos modelos semânticos. Dentre eles, é possível notar modelos para diversas subáreas da saúde, como cenário laboratorial, clínico, cirúrgico e entre outros.

No Brasil, a Troca de Informação para Saúde Suplementar (TISS) é um modelo padrão para troca de informações entre os agentes de saúde suplementar e planos de saúde, que tem por objetivo a uniformização de ações clínicas, administrativas e

financeiras, com a permissão do acompanhamento financeiro das operadoras de convênios médicos. O padrão *Fast Healthcare Interoperability Resources* (FHIR), desenvolvido pela *Health Level 7 International* (HL7®), é um protocolo internacional para envio e recebimento de dados na área da saúde, que contempla informações clínicas e administrativas e está de acordo com a necessidade de integração de dados na área da saúde para otimizar a pesquisa e o desenvolvimento, como afirma Noumeir (2019).

Considerando que os dados de cuidados de saúde primários são a fonte mais rica de dados de saúde (Thiru *et al.*, 2003), ao utilizar conceitos da Ciência da Informação (CI) para realizar a modelagem de arquitetura da informação, busca-se obter a integração de dados e equivalência semântica de diversas fontes heterogêneas, dessa forma, assegurando a fidedignidade da informação, simplificando e unificando a pesquisa e recuperação das informações. Com isso, no Brasil, é necessário enfatizar que há uma preocupação com a interoperabilidade de sistemas médicos, no qual evidenciou-se por meio da portaria nº 2.073, de 2011, do Ministério da Saúde, sendo uma das recomendações desta adotar ontologias e terminologias para lidar com as questões de interoperabilidade de Sistemas de Informação (Brasil, 2011).

1.1 Problema de Pesquisa

Com base na importância do cuidado de saúde e na velocidade da evolução tecnológica, a troca de informações entre diferentes sistemas de saúde é mais que uma necessidade; é um requisito para o avanço da prestação de cuidados de saúde de maneira eficiente. A integração de tecnologias de informação no setor tem desempenhado um papel importante, não apenas na melhoria da gestão de dados de saúde, mas também na facilitação do acesso a informações críticas para a tomada de decisões clínicas. Contudo, esses avanços trazem desafios, especialmente no que diz respeito à gestão eficiente e ao uso estratégico dos dados acumulados no decorrer da vida dos pacientes.

A realidade operacional de muitos sistemas de saúde revela o problema central da pesquisa: a dificuldade em estabelecer uma troca de informações fluida e eficaz entre diferentes entidades de saúde. Esta dificuldade é agravada pela falta de padrões unificados de dados, que resulta em um cenário onde informações médicas são

frequentemente armazenadas de forma não estruturada e fragmentada. Este cenário dificulta o compartilhamento de informações médicas e limita a capacidade dos profissionais de saúde de acessar e utilizar plenamente os dados disponíveis para melhorar a qualidade do atendimento ao paciente.

Neste contexto, a interoperabilidade, que se refere à capacidade dos sistemas de trocar e utilizar informações de maneira eficiente, surge como um elemento habilitador. Este estudo se propõe a explorar a questão da interoperabilidade no setor de saúde, enfrentando, portanto, o desafio de extrair informações valiosas tanto de dados estruturados quanto não estruturados, em diversos cenários, e a dificuldade de desenvolver técnicas para mapear e extrair esses dados de forma eficiente, garantindo uma semântica clara e compartilhável.

1.2 Justificativa

A evolução na prestação de serviços de saúde se tornou uma busca incessante diante do crescimento da demanda e da constante limitação de recursos. Entre as tecnologias utilizadas nesta evolução, destaca-se a tecnologia da informação, em especial, o uso de registros eletrônicos de saúde, que tem o potencial de reformular a coleta, armazenamento e utilização dos dados dos pacientes.

Apesar das inúmeras possibilidades trazidas pelos registros eletrônicos de saúde, a troca eficaz e fluida de informações entre diferentes sistemas de saúde é limitada frequentemente pela ausência de padrões universais de representação de dados. Esta lacuna na interoperabilidade impacta diretamente na agilidade e na qualidade dos serviços prestados, tornando necessário o desenvolvimento de soluções eficazes para essa questão.

O presente estudo justifica-se em várias frentes: primeiro ele proporciona um panorama contemporâneo dos desafios e práticas na arquitetura da informação no domínio da saúde. Adicionalmente, o desenvolvimento de um modelo informacional proposto neste estudo promete ser um instrumento útil para profissionais que lidam com registros eletrônicos de saúde e buscam melhorar sua interoperabilidade.

Ainda, a pesquisa possui a capacidade de elucidar perspectivas relevantes para a elaboração de novas estratégias e políticas de gestão da informação em saúde, objetivando a promoção da interoperabilidade. Esse processo pode resultar na

melhoria do cuidado ao paciente, o que demonstra a aplicabilidade e relevância do estudo.

O papel da Ciência da Informação na melhoria da representação e recuperação de informações nos sistemas de saúde fica evidente pelo uso de técnicas e métodos de tratamento, organização e recuperação dos dados. O estudo proposto buscou, dessa forma, abordar a questão da interoperabilidade, com expectativa de oferecer contribuições para o campo e beneficiar a troca de informações no setor de saúde.

1.3 Metodologia

Os procedimentos metodológicos deste estudo consistem em uma pesquisa básica, pois objetiva “gerar conhecimentos para o avanço da ciência sem aplicação prática prevista” (Prodanov; Freitas, 2013, p.51). Do ponto de vista dos objetivos estabelecidos, trata-se de uma pesquisa exploratória e documental, com a finalidade de proporcionar mais informações sobre o assunto a ser investigado, possibilitando sua definição e seu delineamento, e uso de documentos para o desenvolvimento do trabalho.

Quanto aos procedimentos técnicos empregados, este estudo compõe-se de uma pesquisa bibliográfica, porque se baseia na revisão de um conjunto relevante de fontes bibliográficas, como livros, artigos científicos, teses e dissertações, além de observações e análises de documentos com dados de alergia anonimizados advindos de prontuários eletrônicos, fornecidos pelo Hospital Sírio Libanês (HSL).

A construção da arquitetura do modelo proposto foi iniciada com uma análise detalhada dos dados de alergia fornecidos pelo HSL. Este processo de análise e mapeamento de dados teve como intuito identificar padrões, variações e possíveis lacunas nas informações disponíveis, para posteriormente mapear estes dados conforme o padrão FHIR, garantindo, nesse sentido, a precisão e integridade das informações durante sua transferência.

Com base na análise e mapeamento realizados, foi desenvolvido um modelo de interoperabilidade, utilizando o padrão FHIR com terminologia *Systematized Nomenclature of Medicine - Clinical Terms* (SNOMED CT), construído com a intenção de ser flexível e permitir sua implementação em variados contextos da área da saúde. Para consolidar o estudo, uma implementação do modelo, juntamente com um servidor FHIR, foi configurado, possibilitando em realizar testes e validar protocolos

do padrão. Este modelo, nomeado FHIR-FLOW, foi avaliado com base em sua eficácia em promover a interoperabilidade com FHIR e superar barreiras existentes na troca de informações de saúde, considerando tanto bases de dados estruturadas como não estruturadas.

Com relação às questões teóricas desta pesquisa e aos detalhes dos procedimentos metodológicos aplicados, foi realizada inicialmente uma composição do *corpus* teórico. Este processo foi conduzido por meio de uma pesquisa bibliográfica abrangente, consultando diversas bases de dados relevantes para as áreas da Ciência da Informação e Ciência da Computação. As bases de dados incluíram: Base de Dados em Ciência da Informação (BRAPCI), *Institute of Electrical and Electronics Engineers* (IEEE) *XPlore* e entre outras. Os termos utilizados nas buscas foram "interoperabilidade saúde" e "*health interoperability*", respectivamente. O objeto de análise considerado neste estudo são os artigos científicos.

A partir das buscas realizadas, um total de 29 artigos foram encontrados na BRAPCI, dos quais 8 foram identificados como alinhados à proposta desta pesquisa. Da mesma forma, na base de dados IEEE *XPlore*, dos 17 artigos retornados, 4 foram pertinentes à proposta da pesquisa.

Com a literatura recuperada, os artigos selecionados foram submetidos a uma leitura detalhada, buscando destacar o problema de pesquisa, os objetivos, a metodologia, os resultados e as conclusões de cada trabalho. Este processo permitiu uma compreensão mais profunda dos desafios e avanços recentes no campo da interoperabilidade na saúde.

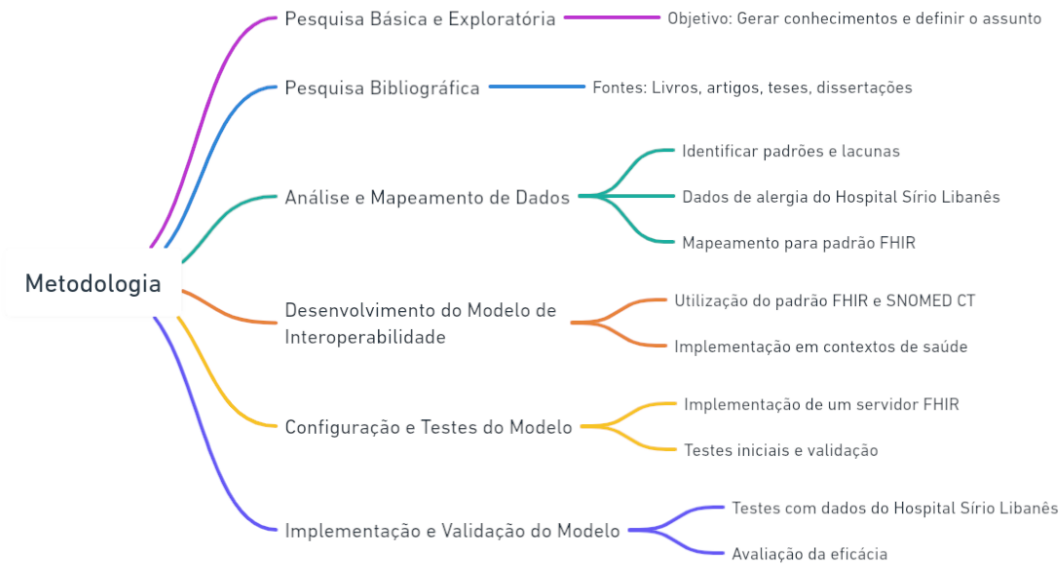
No aspecto empírico desta pesquisa, foi conduzido um estudo sobre o padrão de interoperabilidade de dados FHIR. Esta investigação não se limitou a uma análise teórica da estrutura e composição do padrão, mas também incluiu a implementação concreta de um servidor FHIR. Em paralelo, foram aplicadas técnicas de *Machine Learning* (ML) e Processamento de Linguagem Natural (PLN) para a extração de dados dos documentos. Dada a heterogeneidade dos dados, que podem se apresentar de forma estruturada ou não, o ML e o PLN foram instrumentais na identificação e extração de entidades, permitindo uma compreensão do conteúdo dos dados clínicos. Ademais, foram analisados dados de alergia de prontuários eletrônicos fornecidos pelo Hospital Sírio Libanês, através de uma metodologia de arquitetura da informação para extração de metadados de negócio, que contribuiu estabelecendo

uma conexão entre teoria e realidade do setor de saúde, e permitiu a validação efetiva do modelo de interoperabilidade proposto.

Este estudo também avançou na implementação e validação do modelo de interoperabilidade desenvolvido, fase que envolveu a construção e aplicação do modelo em um ambiente controlado, utilizando dados reais para avaliar sua eficácia. Testes foram conduzidos para assegurar a integração e comunicação eficaz comparando o resultado com a análise que foi realizada com os dados do HSL.

No que diz respeito às limitações deste estudo, a pesquisa concentrou-se apenas na interoperabilidade de dados de saúde existentes e nos desafios associados, sem se aprofundar nas etapas anteriores da construção de novas bases de dados, e também com relação a aplicação terminológica, onde foi considerada somente a terminologia SNOMED CT. O esquema dos passos desenvolvidos na metodologia pode ser observado na Figura 1.

Figura 1 – Esquema do passo a passo da metodologia.



Fonte: Elaborado pelo autor.

A Figura 1 mostra uma visão geral da metodologia da pesquisa, iniciando com pesquisa básica e exploratória, seguida de análise bibliográfica. Posteriormente, ocorre a análise e mapeamento de dados de alergia do HSL, adaptando-os ao padrão FHIR. O desenvolvimento do modelo de interoperabilidade envolve a utilização do padrão FHIR e SNOMED CT, com a seguinte configuração e testes em um servidor FHIR, resultando na implementação e validação do modelo com testes práticos.

1.4 Estrutura da Pesquisa

A seção de introdução tem como objetivo apresentar e delimitar o escopo da pesquisa e expor os principais desafios e oportunidades relacionados ao tema em questão. No texto é detalhado o problema de pesquisa, a justificativa e os objetivos gerais e específicos do estudo. Além disso, é apresentada a metodologia aplicada para o desenvolvimento da pesquisa, bem como a estrutura que será adotada.

Na seção 2, denominada “Panorama dos dados de Saúde”, são abordados os aspectos fundamentais dos dados clínicos de pacientes e discutida a importância desses dados desde o início do contato assistencial, abordando a necessidade de identificação correta da patologia e a escolha assertiva do tratamento. São mencionadas as informações contidas nos prontuários eletrônicos, incluindo dados de origem exclusiva dos pacientes e dados provenientes de contatos assistenciais. Destacam-se também na seção a extensão dos dados de saúde, e suas implicações no cuidado ao paciente, e explorado o conceito de interoperabilidade de dados de saúde, e sua importância na troca de informações entre diferentes sistemas. Apresentam-se os benefícios da interoperabilidade, como a troca de informações na gestão de consultórios, clínicas e hospitais, o compartilhamento seguro de dados do prontuário eletrônico do paciente e a disponibilização ágil de resultados de exames laboratoriais. São abordadas as atividades e recursos envolvidos no processo de interoperabilidade, como a heterogeneidade das fontes de dados e os modelos de troca de informação entre instituições de saúde.

Na seção 3, intitulada “Interoperabilidade de dados de Saúde: Trabalhos Correlatos”, são apresentados estudos correlatos que abordam a integração de dados de saúde. Nessa seção, são mencionadas pesquisas consideradas relevantes, como o trabalho de Petry *et al.* (2008), que propõe um modelo de interoperabilidade utilizando o padrão HL7, e o estudo de Roehrs *et al.* (2018), que apresenta um modelo de integração de dados de saúde a partir de uma base de dados com registros médicos processados. Nesse sentido, também é referido o trabalho de Braunstein (2018), que discute os níveis de interoperabilidade desejados no contexto da saúde. Por fim, são destacados os principais resultados e limitações desses estudos. Ao final da seção são apresentadas as principais diferenças entre os trabalhos correlatos e o trabalho atual.

Na seção 4, denominada “Padrão FHIR”, é discorrida em maior detalhe o padrão FHIR, promovido pela HL7, e seu objetivo de determinar uma transferência representacional do estado para representar as entidades e procedimentos de saúde como recursos. Na seção, é discutido o papel dos recursos no FHIR e como eles definem a estrutura e o conteúdo de informações transmitidas entre sistemas. Menciona-se também a utilização de terminologias no FHIR, que vinculam os dados a vocabulários comuns, como SNOMED, *Logical Observation Identifiers Names and Codes* (LOINC) e *International Classification of Diseases* (ICD). Por fim, é destacado o impacto do uso do padrão FHIR na melhoria do acesso à informação e na qualidade do atendimento ao paciente.

Na seção 5, denominada “*Machine Learning* na Área da Saúde”, é discutido o papel do *Machine Learning,* ou “aprendizado de máquina” na área da saúde. É abordada a necessidade de extração e normalização de dados a partir de documentos médicos, e como o ML pode auxiliar nesse processo, e explorado o uso de algoritmos de ML na categorização de informações em registros médicos e na conversão de dados não estruturados em dados estruturados. Nesta seção, são apresentadas aplicações do ML na medicina, como o diagnóstico é auxiliado por computador, a sua personalização de tratamentos médicos e os desafios e oportunidades do uso de ML na saúde.

Na seção 6, intitulada “Dados de Alergia do Hospital Sírio Libanês: análise com Metadados de Negócio", é elucidada a aplicação prática de integração de dados e interoperabilidade no contexto do HSL. A seção apresenta a importância dos metadados no ambiente de negócios e como eles contribuíram para assegurar a informação correta. Esta inserção proporciona um estudo de caso real, que demonstra as teorias e práticas discutidas nas seções anteriores e evoca *insights* para a construção do modelo de interoperabilidade com FHIR.

Na seção 7, denominada "FHIR-FLOW: Modelo de Mapeamento de Dados de Saúde para FHIR", é apresentado o FHIR-FLOW, um modelo de interoperabilidade cujo objetivo principal é abordar e solucionar os desafios de interoperabilidade de dados ao agir como uma ponte, transformando dados brutos e heterogêneos em informações padronizadas e prontas para interoperabilidade. A seção explora a concepção e funcionalidade do modelo, demonstrando como ele pode ser aplicado em diferentes contextos da saúde, além da prova de conceito com os dados do HSL.

2 PANORAMA DOS DADOS DE SAÚDE

Os dados clínicos de pacientes são fundamentais desde o início do contato assistencial, a começar pela identificação correta da patologia, que possibilita a eleição adequada de tratamento, medicações e procedimentos. A escolha correta do tratamento pode não só diminuir o tempo de duração da patologia, como interferir diretamente na prevenção de óbitos, dado que “o erro de diagnóstico pode ser a maior preocupação de segurança do paciente não tratada nos Estados Unidos, responsável por cerca de 40.000 a 80.000 mortes anualmente”, como afirma Graber (2017, 1).

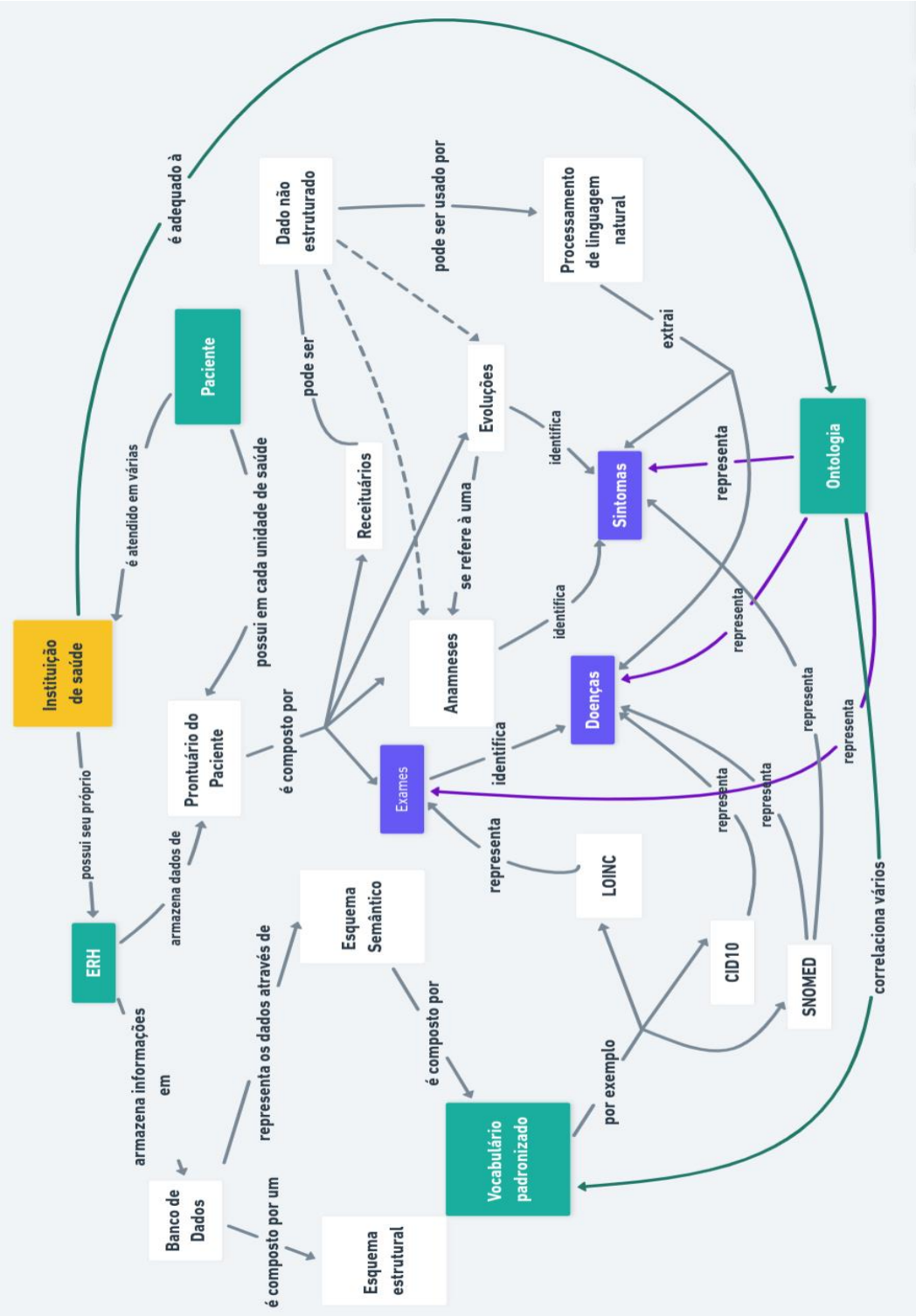
A extensão dos dados de saúde contempla toda a abrangência de dados clínicos, desde informações de origem exclusiva dos pacientes, como tipo sanguíneo e etnia, até os dados provenientes de contatos assistenciais, como resultado de exames, anamneses, evoluções e receituários. Para ajudar na etapa de identificação e monitoramento das patologias são utilizados resultados de exames, que subsidiam os profissionais nas tomadas de decisões de diagnósticos quanto nas evoluções posteriores, mediante a medicações e procedimentos. Atualmente, de acordo com Graber (2017), existem mais de 4.000 testes de laboratório selecionáveis, e um número comparativamente desconcertante de opções de imagem.

Além dos dados puramente clínicos, informações geográficas desempenham um papel importante, especialmente, quando se trata de identificar fatores de natureza epidemiológica. Hung (2020) ressalta que a natureza detalhada dos dados do Sistema de Informação em Saúde, combinada à sua capacidade de se relacionar com outros dados geográficos, pode impulsionar pesquisas significativas, aprofundando a compreensão sobre a epidemiologia das doenças.

Para proporcionar uma compreensão clara e detalhada das informações e suas inter-relações no contexto da saúde, é fundamental que ocorra a representação desses dados através de mapas conceituais. Essas ferramentas destacam-se na representação de conceitos clínicos e assistenciais, evidenciando a complexidade das inter-relações entre dados clínicos. Ademais, a inclusão de ontologias, que são usadas para estruturar e codificar adequadamente esses dados, é importante. Segundo Pickler (2007), as ontologias são essenciais na *Web* Semântica e desempenham um papel vital na representação do conhecimento e na contextualização dos dados, pois possibilitam realizar interpretação semântica das informações por máquinas e sistemas, facilitando a integração de dados entre

diferentes plataformas. No setor da saúde, sistemas como LOINC, ICD10 e SNOMED são exemplos da aplicação de ontologias para a codificação padronizada de dados.

Figura 2 - Mapa conceitual do ambiente informacional de prontuário eletrônico



Fonte: Adaptado de IQBAL *et al.*, 2011.

Na figura 2, pode-se observar um sistema de Registros Eletrônicos de Saúde (EHR, tradução livre), no qual como diversos componentes interagem de forma integrada. No coração deste sistema está o próprio EHR, que atua como um núcleo central armazenando e gerenciando informações essenciais de saúde.

O EHR está intrinsecamente ligado a um banco de dados, onde armazena informações dos pacientes, incluindo seus prontuários, que, por sua vez, são uma compilação abrangente de dados do paciente, contendo receituários, exames, anamneses e evoluções. Ele é estruturado em dois aspectos principais: esquema estrutural, que define a arquitetura dos dados; e esquema semântico, que interpreta o significado desses dados. As instituições de saúde, cada uma com seu próprio sistema de EHR, são conectadas ao núcleo, refletindo a realidade de que os pacientes são atendidos em várias instituições, e, cada uma delas, contribui para o acervo de dados do paciente, cuja construção ocorre de forma distribuída, ao longo do tempo.

Dentro do esquema semântico, há um papel que é desempenhado pelo vocabulário padronizado, que inclui terminologias como CID10, LOINC e SNOMED CT. Esses vocabulários são usados para representar exames, doenças e sintomas de maneira padronizada e garantir que os dados sejam compreensíveis e interoperáveis entre diferentes sistemas, facilitando a compreensão e o compartilhamento de informações. Além de que, parte dos dados nos prontuários, como receituários, anamneses e evoluções, são frequentemente não estruturados. Neste cenário, entra em atuação o processamento de linguagem natural, uma tecnologia que extrai informações significativas desses dados não estruturados, como sintomas e diagnósticos de doenças.

A ontologia, outra camada neste sistema, representa exames, sintomas e doenças, auxilia na organização e na interpretação dos dados de saúde, e facilita a identificação de padrões e conexões, como a relação entre exames específicos e as doenças que eles ajudam a diagnosticar. Portanto, neste sistema de EHR, cada componente desempenha um papel importante na gestão dos registros de saúde, trabalhando em conjunto para assegurar que os dados de saúde sejam não apenas armazenados, mas também interpretados e utilizados de maneira que melhore o atendimento ao paciente e a eficiência das instituições de saúde. Em um cenário ideal, permitiria que os dados dos pacientes sejam compartilhados por diferentes instituições de saúde.

2.1 Interoperabilidade de Dados de Saúde

Na área da saúde, a interoperabilidade vem crescendo para permitir a troca de dados entre os diferentes sistemas e ferramentas utilizadas, gerando mais informações valiosas no cuidado do paciente. Nesse sentido, Pine (2019, p. 1) complementa que:

> As pesquisas sobre interoperabilidade e troca de informações entre sistemas de tecnologia da informação destacam o uso de dados para uma variedade de propósitos, incluindo pesquisa, gestão, melhoria da qualidade e prestação de contas.

Dentre os principais benefícios no setor da saúde, a interoperabilidade permite:

- Trocar de informações na gestão de consultórios, clínicas e hospitais. Ela é especialmente útil para instituições que atuam em todos os níveis de atenção ao paciente, permitindo o rastreio clínico do indivíduo nos serviços utilizados;
- Compartilhar dados do Prontuário Eletrônico do Paciente (PEP) com segurança, para aumento de evidências que subsidiam decisões clínicas;
- Disponibilizar resultados de exames laboratoriais e de radiologia, proporcionando emitir e obter laudos com maior agilidade.

A interoperabilidade no setor da saúde, baseada em um modelo abrangente, proporciona uma visão integral da saúde do paciente ao reunir, compartilhar e utilizar diferentes informações clínicas e administrativas. Este modelo aumenta a segurança e eficiência da assistência à saúde, facilitando a comunicação entre os profissionais envolvidos no cuidado do paciente. Isso reduz procedimentos e exames duplicados, o que não somente promove agilidade e transparência na troca de informações, mas também contribui significativamente para a otimização dos processos e a redução de custos no sistema de saúde. Entretanto, as atividades e recursos envolvidos no processo de interoperabilidade são complexos, devido a fatores como a própria heterogeneidade das diversas fontes de dados, que apresentam desde diferenças estruturais e semânticas até os modelos de troca de informação entre as instituições de saúde.

As diferenças estruturais podem ser observadas no modo como os sistemas organizam e armazenam seus dados, como quantidade de tabelas, tipagem de dados e escolha por texto livre ou informações tabuladas. As diferenças semânticas podem ser percebidas desde a escolha de vocabulários adotados por cada instituição, que,

apesar de possuir essencialmente o mesmo significado, são identificados por códigos e descrições divergentes.

Para que as instituições possam trocar informações de forma precisa e automática, os documentos clínicos eletrônicos devem fazer uso de códigos clínicos estabelecidos, também chamados de vocabulários controlados, como de SNOMED CT, LOINC e ICD-10. No entanto, não existe um esquema de codificação universalmente aceito que encapsula todas as informações clínicas, como afirma Hamm (2007). Atualmente, existem diversas propostas de soluções e caminhos a serem adotados, para realizar a interoperabilidade que se complementam, e para padrões trocar informações, como o TISS e o padrão FHIR.

Mesmo considerando os desafios existentes, a adoção de interoperabilidade de dados de saúde é uma prática eficiente, pois traz às instituições a possibilidade de oferecer um tratamento mais eficiente ao paciente, ao mesmo tempo que otimiza a utilização de seus recursos, aumenta a eficiência nos processos, e reduz custos. Na próxima seção são apresentados trabalhos correlatos realizados no cenário de interoperabilidade de dados de saúde, com o intuito de expor suas principais características e desafios.

3 INTEROPERABILIDADE DE DADOS DE SAÚDE: TRABALHOS CORRELATOS

Para a elaboração desta revisão de literatura, foram analisados diversos artigos científicos que abordaram a interoperabilidade de dados de saúde em sistemas de prontuários eletrônicos do paciente, com ênfase no uso do padrão.

A seleção dos artigos foi baseada em critérios de relevância, priorizando pesquisas acadêmicas que investigaram o uso de técnicas contemporâneas de mapeamento semântico, *Machine Learning* e outras abordagens, visando a interoperabilidade dos dados de saúde nos prontuários eletrônicos. Cada artigo selecionado foi submetido a uma análise, considerando essas categorias, e suas respectivas implicações, para promoção da interoperabilidade de dados de saúde nos sistemas de prontuários eletrônicos, com foco especial na aplicação do padrão FHIR, como uma solução viável.

O estudo de Petry *et al.* (2008) introduziu um modelo baseado no padrão HL7, projetado para um servidor de troca de mensagens, que envolve 77 municípios da Rede Catarinense de Telemedicina (RCTM). Esta abordagem resultou em reduções expressivas de custos e tempo de atualização, e ofereceu um reforço notável na segurança do Portal de Telemedicina. Contudo, a abordagem de mapeamento dos dados originais permaneceu obscura, não disponibilizados detalhes no estudo.

Roehrs *et al.* (2018) apresentaram um modelo de integração de dados de saúde vindos de registros médicos de 38.645 pacientes adultos. Para tal, utilizaram padrões renomados, incluindo *open*EHR, HL7 FHIR e *Medical Information Mart for Intensive Care* (MIMIC-III). Um dos triunfos desse trabalho foi a implementação eficaz de técnicas de Inteligência Artificial (IA) e processamento de linguagem natural para impulsionar a interoperabilidade. No entanto, uma limitação saliente foi a concentração exclusiva em dados já padronizados, sem atenção a dados brutos ou não conformes.

Enquanto Braunstein (2018) abordou os níveis de interoperabilidade desejados e discutiu as complexidades e desafios associados ao padrão anterior da HL7, valorizando o padrão FHIR como uma solução promissora e evidenciando sua adoção por instituições de grande envergadura, como *Medicare* e *Veteran's Administration* (VA). Porém, embora sua discussão tenha sido rica em *insights*, Braunstein (2018) não delineou um modelo ou fluxo específico para a adaptação de dados brutos ao padrão FHIR.

Chatterjee *et al.* (2022) focaram na problemática da heterogeneidade na armazenagem e troca de dados em sistemas de informação de saúde digital. Eles propuseram utilizar o padrão FHIR juntamente com o SNOMED CT, para conectar dados de saúde pessoais a prontuários eletrônicos de saúde e, como prova de conceito, desenvolveram o aplicativo de coaching de saúde, denominado como *eCoach*. A combinação eficaz de HL7 FHIR e vocabulários SNOMED CT, bem como a implementação de padrões de qualidade de interoperabilidade, foram pontos fortes destacados. Contudo, as técnicas empregadas ainda não foram testadas ou validadas em ambientes clínicos reais, com uma variedade mais ampla de dados e situações, o que sugere que a generalização para cenários de saúde mais complexos pode ser um desafio.

Na pesquisa de Balch *et al.* (2023), exploraram a aplicação e potencial dos sistemas de informação clínica habilitados para *Machine Learning* o contexto da transformação da entrega e pesquisa em saúde. Sublinharam a crescente integração do padrão de dados *Fast Healthcare Interoperability Resources* nesses sistemas, apesar das variações em métodos de implementação. A pesquisa revelou avanços notáveis, como uso inovador de sistemas em nuvem, redes Bayesianas, estratégias de visualização, técnicas de conversão de dados não estruturados para FHIR. Entretanto, identificaram-se limitações significativas, como enfrentamento de barreiras de interoperabilidade, com prontuários eletrônicos de saúde em sistemas avançados e a presença de carência de evidências, externamente validadas quanto à sua eficácia clínica.

Pimenta *et al.* (2023) investigaram a contínua problemática da interoperabilidade dos dados clínicos, mesmo diante dos avanços tecnológicos na área da saúde. Enfatizaram o padrão FHIR como uma ferramenta construída sobre padrões da *web*, destacando sua flexibilidade, facilidade de uso e implementação. A pesquisa teve como objetivo avaliar o potencial do FHIR, identificando componentes essenciais que podem otimizar a interoperabilidade e estabelecer uma conexão mais eficiente entre sistemas de saúde e fontes de dados clínicos. Ao adotar o FHIR, observou-se que os profissionais de saúde poderiam comunicar-se de maneira mais aprimorada. Embora o trabalho tenha abordado a promissora implementação do padrão FHIR para interoperabilidade de dados clínicos, ele baseou-se principalmente em uma revisão da literatura já existente, não tendo descobertas e conclusões validadas experimentalmente em ambientes práticos, o que limita a aplicabilidade

direta das recomendações em cenários clínicos reais.

A seguir, na Tabela 1, constam os dados comparativos entre os trabalhos estudados:

Tabela 1 – Dados comparativos entre trabalhos correlatos.

Autor(es) e Ano	**Padrões Abordados**	**Pontos Fortes**	**Limitações**	**Uso de IA?**
Petry *et al.* (2008)	HL7	Redução de custos e tempo, Segurança no Portal de Telemedicina	Falta de detalhes sobre a abordagem de mapeamento dos dados originais para o padrão HL7	Não
Roehrs *et al.* (2018)	*open*EHR, HL7 FHIR, MIMIC-III	Uso eficaz de IA e NLP	Concentração em dados já padronizados	Sim
Braunstein (2018)	HL7, FHIR	Valorização do padrão FHIR	Não delineou modelo para adaptação de dados brutos ao FHIR	Não
Chatterjee *et al.* (2022)	FHIR, SNOMED CT	Combinação de HL7 FHIR e SNOMED CT	Não testado em ambientes clínicos reais	Sim
Balch *et al.* (2023)	FHIR	Uso de sistemas em nuvem, conversão de dados não estruturados para FHIR	Barreiras de interoperabilidade com prontuários eletrônicos, falta de evidências validadas	Sim
Pimenta *et al.* (2023)	FHIR	Flexibilidade e facilidade do FHIR	Baseado em revisão da literatura, sem validação experimental	Não
Trabalho Atual	FHIR	Uso de SNOMED CT e proposta de modelo generalista para mapeamento para padrão FHIR	Restrição a SNOMED CT, e testado somente com dados de alergia	Sim

Fonte: Elaborado pelo autor.

4 PADRÃO FHIR

O avanço da tecnologia da informação tem desempenhado um papel significativo na transformação dos sistemas de saúde, proporcionando melhorias na coleta, armazenamento e troca de informações médicas. No entanto, a interoperabilidade eficiente e segura dos dados de saúde continua sendo um desafio enfrentado pela indústria e pelos profissionais de saúde.

Nesse contexto, o padrão FHIR, foi criado com objetivo de fornecer um conjunto de recursos e *Application Programming Interface* (APIs) baseadas em *Representational State Transfer* (REST), para facilitar o acesso e uso de dados de saúde do paciente, em um nível granular (Ayaz *et al.*, 2021), e representar as entidades e procedimentos de saúde como recursos, como por exemplo, paciente, medicação, observação, e notas clínicas. A ideia central consiste em construir um conjunto básico de recursos que, isoladamente ou combinados, atende a muitos usos comuns casos (Stan; Miclea, 2018). A proposta do FHIR é oferecer uma abordagem atualizada para o compartilhamento de informações de saúde, superando as limitações dos padrões anteriores.

Uma das principais vantagens do FHIR está em sua capacidade de fornecer uma camada semântica, permitindo associação dos dados de saúde às terminologias internacionais, como SNOMED, LOINC e ICD. Esse processo facilita a padronização e interoperabilidade dos dados, garantindo que as informações sejam compreendidas e interpretadas corretamente pelos sistemas de saúde. Além disso, o FHIR adota uma abordagem orientada a serviços, oferecendo funcionalidades como consultas, serviços de notificação e autenticação, que contribuem para a troca segura e eficiente de informações entre sistemas de saúde.

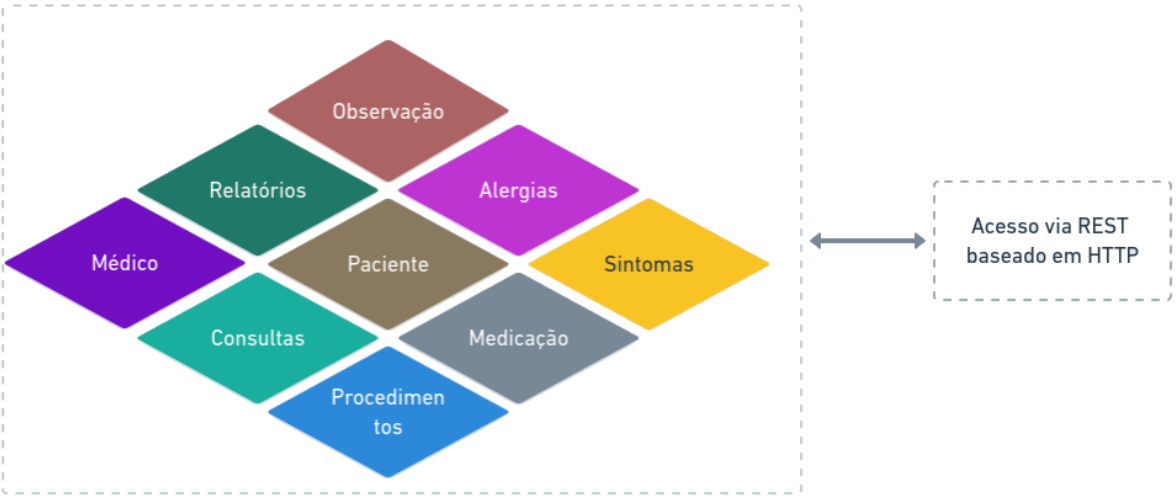
4.1 Arquitetura Geral do FHIR

A arquitetura do FHIR segue os princípios do estilo arquitetural REST, no qual é orientada em torno de recursos, que são a unidade central do FHIR. Esses recursos representam diversas entidades e conceitos na área da saúde, como pacientes, medicamentos, condições clínicas, agendamentos, procedimentos, entre outros (Saripalle *et al.*, 2019). Cada recurso é definido em formatos de dados estruturados, como em *Extensible Markup Language* (XML), *JavaScript Object Notation* (JSON) ou

Resource Description Framework (RDF), e são acessíveis através de APIs baseadas em *Hypertext Transfer Protocol* (HTTP). Essa abordagem modular, que expõe entidades de dados de saúde, como serviços acessíveis via APIs baseadas em HTTP, diferencia o FHIR de outros padrões mais centrados em documentos, tendo a sua implementação facilitada pela flexibilidade na escolha do formato de representação de dados (Saripalle *et al.*, 2019).

Além dos recursos, a arquitetura do FHIR também incorpora perfis, que são definições adicionais e restrições usadas para adaptar os recursos aos contextos específicos de implementação. Esses perfis permitem a personalização e extensão dos recursos, garantindo a conformidade com requisitos e necessidades locais.

Figura 3 – Visão de acesso aos recursos na arquitetura FHIR



Fonte: Adaptada de HL7 FHIR, 2021.

A Figura 3 representa o acesso aos recursos FHIR, que são considerados componentes essenciais dos dados de saúde. O quadrado maior representa esses recursos, como Observação, Relatórios, Paciente, Médico, entre outros. O quadrado "Acesso" simboliza a maneira pela qual os usuários e sistemas interagem com esses recursos, usando APIs baseadas em HTTP/REST. A figura visualiza a relação entre os recursos FHIR e o acesso a eles, destacando a importância do padrão FHIR na disponibilização dos dados de saúde e promovendo a interoperabilidade entre os sistemas.

4.2 Recursos FHIR

O FHIR utiliza um componente central denominado *Resource*, que é essencial na definição da estrutura e conteúdo das informações trocadas entre diferentes sistemas de saúde. Um recurso, no FHIR, pode ser estruturado por meio de composição, o que significa que ele pode incluir referências a outros recursos dentro do sistema.

Todos os recursos no FHIR apresentam uma forma comum de representação. Eles utilizam tipos de dados primitivos, como *integer* (número inteiro), *string* (cadeia de caracteres) e *boolean* (valor verdadeiro ou falso), para garantir a consistência na troca de informações. Além disso, estes recursos podem estar associados a outros tipos de recursos específicos, como *Patient* (Paciente), *Visit* (Visita) e *Drug* (Medicamento).

Um aspecto importante dos recursos no FHIR é a inclusão de uma parte legível por humanos, chamada *human-readable* que consiste em uma narrativa livre, e permite que informações adicionais, que podem não ser facilmente representadas pelos tipos de dados primitivos, sejam incluídas, garantindo assim uma compreensão mais ampla do contexto e do conteúdo dos dados transmitidos.

Figura 4 – Diagrama UML parcial do recurso FHIR *Patient*

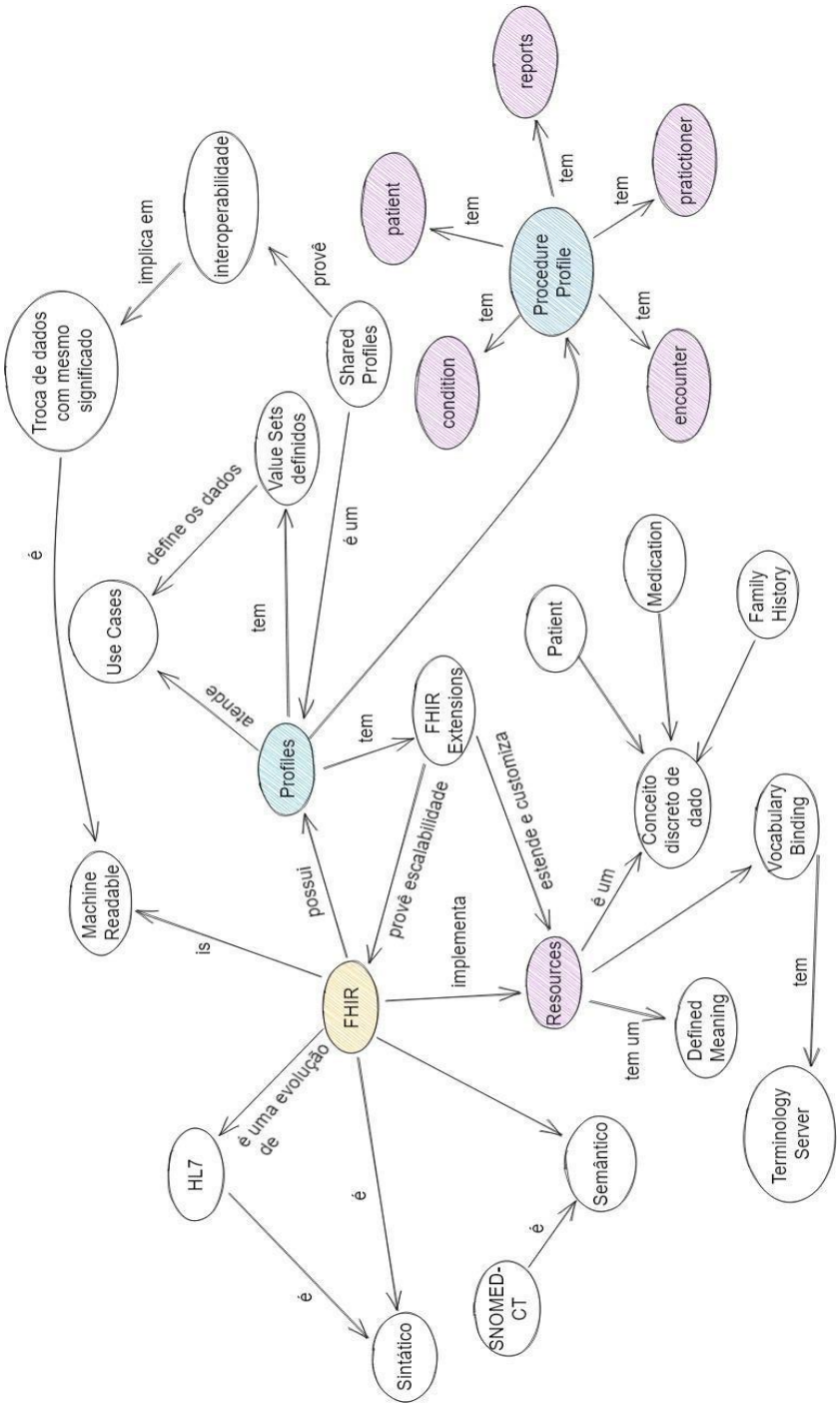
Fonte: Adaptada de HL7 FHIR, 2021.

A Figura 4 apresenta um diagrama *Unified Modeling Language* (UML) representando o recurso *Patient*, baseada no recurso de mesmo nome do FHIR. O recurso inclui respectivamente atributos de identificador, *status* de atividade, nome, celular, data de nascimento, endereço e foto.

Outros componentes localizados na arquitetura FHIR são os *ValueSets*, que são usados para padronização e reutilização de dados em cenários clínicos. Os *ValueSets* definem um conjunto de valores aceitos para um cenário de saúde específico, agregando valor semântico aos dados e facilitando a interoperabilidade.

Além dos *ValueSets*, o FHIR incorpora o componente *Terminologies*. Este componente é responsável por vincular dados clínicos às terminologias internacionais já estabelecidas, como SNOMED, LOINC, ICD-9 e ICD-10. Essa integração permite o reaproveitamento de vocabulários existentes e amplamente utilizados, e possibilita a realização de mapeamentos entre diferentes terminologias com base em evidências. Portanto, o FHIR, por meio dos *ValueSets* e do componente *Terminologies*, proporciona uma estrutura para a interoperabilidade de dados em saúde, permitindo a padronização de vocabulários e valores em cenários clínicos, além de facilitar a integração com terminologias internacionais, o que viabiliza a precisão na troca de informações de saúde.

Figura 5 – Mapa conceitual da arquitetura do padrão FHIR.



Fonte: Adaptada de HL7 FHIR, 2021.

Na Figura 5 pode-se observar a arquitetura geral do FHIR, que representa uma evolução do padrão HL7 e é considerado uma tentativa de melhorar a interoperabilidade no setor de saúde. Ao contrário do HL7, que é predominantemente sintático, o FHIR é tanto sintático quanto semântico, o que significa que ele aborda

não apenas a estrutura dos dados, mas também o seu significado. O FHIR é composto por *Resources*, que são conceitos discretos de dados. Estes *Resources* possuem um significado definido e incluem exemplos como *Vocabulary Binding, Patient, Medication* e *Family History*, que são recursos que, por sua vez, representam informações do contato assistencial.

Para garantir flexibilidade e escalabilidade, o FHIR permite *Extensions,* que são artefatos que podem estender e personalizar os *Resources*. Os *Profiles* no FHIR são uma maneira de definir regras específicas e restrições para os *Resources*. Eles contêm *Extensions* do FHIR, atendem a diferentes cenários e possuem *Value Sets* definidos, que são conjuntos finitos de valores aceitos em um determinado dado.

O *Procedure Profile* é um exemplo de um *Profile*, que abrange vários conceitos como *Condition, Encounter, Practitioner, Reports e Patient*. Há também *Shared Profiles*, que são projetados para promover a interoperabilidade, garantindo que os dados sejam trocados com o mesmo significado entre os sistemas, através de compartilhamento de definições FHIR entre instituições. O FHIR enfatiza a troca de dados com o mesmo significado, e, sendo *Machine Readable*, garante que as máquinas possam ler, interpretar e usar esses dados de forma eficiente. Dessa forma, o padrão FHIR representa uma evolução significativa na busca pela interoperabilidade no setor de saúde, abordando tanto aspectos sintáticos quanto semânticos dos dados e proporcionando flexibilidade através de *Resources*, *Extensions* e *Profiles*.

Neste capítulo destacou-se a importância do FHIR para a interoperabilidade na saúde. No entanto, a simples conexão entre sistemas representa somente uma parte do potencial. A análise profunda desses dados unificados, principalmente, os não estruturados, é uma demanda latente. No próximo capítulo, discute-se o papel do *Machine Learning* neste cenário. Com sua capacidade de lidar com grandes volumes de dados e extrair insights valiosos, pois ele potencializa a análise clínica e auxilia na própria questão da interoperabilidade, facilitando a conversão e integração de dados.

5 MACHINE LEARNING NA ÁREA DA SAÚDE

Na medicina atual, a extração e normalização de dados, a partir de documentos médicos, são desafios significativos, especialmente em clínicas menores ou em regiões com transição lenta para sistemas digitais, devido às limitações financeiras, falta de treinamento ou infraestrutura tecnológica.

Profissionais de saúde, frequentemente, dependem de registros médicos em papel, onde gerenciamento, recuperação e acesso são trabalhosos, lentos e propensos a erros. Essa ineficiência aumenta o risco de erros clínicos e impacta a qualidade do atendimento ao paciente. Além disso, a ausência de padrões unificados, para representação e armazenamento de dados de saúde, agrava os problemas de interoperabilidade, criando barreiras a comunicação eficiente entre diferentes sistemas, dispositivos e plataformas.

Neste cenário, o *Machine Learning* surge como uma ferramenta transformadora, no qual é considerado um campo da Inteligência Artificial que envolve o desenvolvimento de algoritmos capazes de aprender e fazer previsões ou decisões baseadas em dados (Ghassemi *et al.*, 2020). O ML permite que sistemas melhorem automaticamente sua performance com experiência, sem serem explicitamente programados para cada situação específica. Desta maneira, na saúde, o *Machine Learning* pode ser aplicado para melhorar a eficiência operacional em uma série de trabalhos, incluindo conversão de registros médicos em papel para formatos digitais. Algoritmos de ML, por exemplo, podem ser treinados para reconhecer padrões complexos em dados de saúde. Além disso ele também tem um papel útil na interoperabilidade entre sistemas de saúde, pois técnicas de processamento de linguagem natural, um subcampo do ML, podem extrair e organizar informações de diversas fontes de dados automaticamente.

5.1 Potencial e aplicações

A aplicação de *Machine Learning* na medicina tem gerado inovações significativas revolucionado a maneira como a saúde é abordada. Esta transformação é evidente na melhoria da qualidade e eficiência do cuidado ao paciente, especialmente devido ao fato de que os EHRs se tornaram uma fonte de dados extremamente rica, conforme destacado por Ghassemi *et al.* (2020).

O ML, com sua capacidade de aprender e se adaptar continuamente a novos conjuntos de dados, mostra-se útil na extração e normalização de dados médicos

essenciais. Algoritmos de ML são capazes de ser treinados para reconhecer, categorizar e interpretar informações em registros médicos, que é um processo fundamental para a transformação de dados não estruturados em estruturados. Este processo está intimamente ligado à Ciência da Informação e desempenha um papel crucial na recuperação de informações. Além do mais, técnicas avançadas de processamento de linguagem natural permitem que modelos de ML interpretem textos escritos e transcrevam dados médicos em formatos normalizados, viabilizando a interoperabilidade entre diferentes sistemas de saúde e contribuindo para uma gestão mais eficiente dos registros, com uma prática médica mais informada e baseada em evidências.

Essa eficiência e precisão na gestão de dados médicos, através do *Machine Learning*, abre caminho para uma compreensão mais profunda dos diferentes tipos de aprendizado de máquina e suas aplicações específicas na medicina e na Ciência da Informação. A Tabela 2 apresenta os principais tipos de *Machine Learning* e faz uma correlação com conceitos fundamentais da CI.

Tabela 2 – Principais tipos de *Machine Learning* e sua correção com a CI

Tipo de Aprendizado	**Explicação**	**Correlação com a Ciência da Informação**
Aprendizado Supervisionado	O algoritmo aprende a partir de exemplos rotulados. Ele recebe pares de entrada e saída e aprende a mapear um no outro.	Corresponde ao conceito de "indexação" na Ciência da Informação. Assim como na indexação, o aprendizado supervisionado requer um conjunto de dados de treinamento bem definido e rotulado para que o algoritmo possa aprender e fazer previsões precisas.
Aprendizado Não Supervisionado	O algoritmo aprende a encontrar padrões e relações nos dados por conta própria. Não recebe saídas para as entradas.	Relaciona-se ao conceito de "recuperação de informações" na CI. Assim como na recuperação de informações, o aprendizado não supervisionado analisa conjuntos de dados não rotulados para descobrir padrões e estruturas subjacentes.
Aprendizado por Reforço	O algoritmo aprende através de interações e feedback do ambiente. Recebe recompensas ou punições com base em suas ações.	Correlaciona-se ao conceito de "*feedback* do usuário" na Ciência da Informação. Assim como o *feedback* do usuário é usado para ajustar e melhorar a precisão dos sistemas de recuperação de informações, o aprendizado por reforço ajusta seu comportamento com base no *feedback* para melhorar seu desempenho ao longo do tempo.

Fonte: Adaptada de JANIESCH *et al.*, 2021.

Neste cenário, os modelos de ML têm sido aplicados em várias áreas da medicina, desde diagnósticos auxiliados por computador até a criação de planos de tratamento personalizados. No diagnóstico auxiliado por computador, algoritmos de ML são usados para detectar padrões em imagens médicas que podem ser desafiadores para a identificação humana. Os modelos de ML têm sido utilizados para detectar tumores em imagens de ressonância magnética e para identificar doenças oculares em imagens de retina, conforme é destacado por Javaid *et al.* (2022).

O ML também tem um papel útil na personalização de tratamentos médicos, pois algoritmos podem analisar dados de pacientes, incluindo histórico médico, resultados de exames laboratoriais e informações genômicas, para desenvolver planos de tratamento personalizados. O uso do ML pode aumentar a eficácia do tratamento, reduzindo efeitos colaterais e melhorando a qualidade de vida dos pacientes (Javaid *et al.*, 2022).

A seguir, são destacados os principais pilares da aplicação de ML na área da saúde, que variam desde a prevenção e preparação para surtos de doenças até o aprimoramento de tratamentos e diagnósticos médicos. Segundo Janiesch *et al.* (2021), esses pilares incluem:

- Previsão de Surto: utilização de algoritmos para se antecipar e preparar para surtos de doenças, maximizando a prontidão e a resposta.
- Descoberta e Fabricação de Medicamentos: aceleração da identificação de compostos terapêuticos e otimização dos processos de fabricação.
- Modificação Comportamental: uso de dados e análises para incentivar comportamentos saudáveis e mudanças de estilo de vida.
- Registros de Saúde Inteligentes: aprimoramento a gestão e análise dos prontuários eletrônicos para extração de insights clínicos.
- Coleta de Dados Colaborativa: potencializa a coleta de informações de múltiplas fontes, tornando a pesquisa e o diagnóstico mais abrangentes.
- Melhor Radioterapia: otimiza tratamentos de radioterapia para maior eficácia e minimização de efeitos colaterais.
- Diagnóstico de Imagens Médicas: melhora a precisão do diagnóstico por meio da análise automatizada de imagens, como raios-X e ressonância magnética.
- Clínica e Pesquisa: facilita a investigação clínica e a pesquisa médica, através da análise de grandes conjuntos de dados.

Esta representação visual destaca o potencial do Machine Learning na transformação e avanço do setor de saúde, abrangendo desde a prevenção até a intervenção terapêutica.

Por fim, no contexto de interoperabilidade de dados, o uso de ML na medicina também oferece oportunidades para melhorar o acesso à informação e a qualidade do atendimento ao paciente. A aplicação de técnicas de ML, como proposto por Bouh, Hossain e Ahmed (2023), permite extrair e padronizar dados médicos de documentos digitalizados de acordo com os padrões FHIR. Nesse sentido, ele não apenas aprimora a eficiência dos dados de saúde, mas também contribui para a interoperabilidade deles, viabilizando a troca e aplicação de informações em diferentes sistemas.

5.2 Processamento de Linguagem Natural

Dentro da Ciência da Informação, o Processamento de Linguagem Natural é uma área interdisciplinar que integra conhecimentos da Ciência da Computação, Inteligência Artificial e Linguística. Segundo Caseli *et al.* (2022), o PLN é essencial para a CI, fornecendo métodos e ferramentas para compreender, interpretar e gerar linguagem humana. Esta área é fundamental para aprimorar a interação entre computadores e usuários, tornando-a mais natural e intuitiva.

O PLN utiliza algoritmos e modelos matemáticos para analisar a estrutura e o significado da linguagem. Inicialmente, emprega-se modelos estatísticos e de aprendizado profundo para identificar padrões linguísticos. Posteriormente, técnicas de análise sintática e semântica são aplicadas para examinar a gramática e o significado das palavras e frases. O objetivo final do PLN é superar a barreira entre a linguagem natural humana e a linguagem de máquina, permitindo o desenvolvimento de sistemas que processam a linguagem natural de maneira eficaz e intuitiva.

Após a etapa inicial de compreensão dos padrões linguísticos, o PLN conforme descrito por Caseli *et al.* (2022), avança para a análise léxica. Essa fase consiste em decompor o texto em unidades menores, como palavras ou termos, denominados como *tokens*. A análise léxica é essencial para identificar e isolar os elementos básicos da linguagem, e uma vez realizada a *tokenização*, segue-se para análise sintática. Esta etapa envolve determinar as funções gramaticais dos *tokens*, permitindo a construção de árvores sintáticas, que ilustram a estrutura gramatical das frases. Estes

processos são importantes para converter a linguagem natural em um formato estruturado, acessível para o processamento por máquinas.

Após a análise sintática, o PLN avança para análise semântica. Esta fase tem como objetivo entender o significado das palavras e sentenças, que é alcançado ao mapear palavras e frases às entidades em um domínio específico, dessa forma, solucionando ambiguidades de sentido e assegurando uma interpretação precisa do texto. A análise semântica é, portanto, um passo fundamental para garantir que a máquina não apenas entenda a estrutura da linguagem, mas também seu conteúdo e nuances.

Métodos como a Extração de Entidades Nomeadas (NER) e a Análise de Sentimento são exemplos de como o PLN utiliza técnicas avançadas, para identificar e classificar informações em textos. Oliveira *et al.* (2022) destacam o uso de representações vetoriais, como *embeddings* de palavras, que são fundamentais no PLN. Esses *embeddings* são transformações de palavras em vetores numéricos em espaços de alta dimensionalidade, permitindo a captura de relações semânticas e sintáticas. Neste espaço vetorial, palavras semanticamente similares são posicionadas próximas umas das outras, facilitando identificação de sinônimos, antônimos e outras relações semânticas. Esta técnica é essencial para modelos de linguagem avançados como o *Bidirecional Encoder Representations from Transformers* (BERT), que aprendem nuances contextuais e semânticas da linguagem.

Na Ciência da Informação, os *embeddings* de palavras criam uma ponte entre informações textuais não estruturadas e estruturas de dados quantitativas e analíticas. Esta conversão de palavras e frases em representações numéricas torna possível a organização, recuperação e análise de dados textuais de forma eficiente, promovendo a descoberta de conhecimento e a extração de *insights*, a partir de grandes volumes de texto.

Nesse sentido, Patel *et al.* (2021) ressaltam a utilidade do PLN na saúde, especialmente, na transformação digital de registros médicos. O PLN facilita a estruturação de dados não estruturados em formatos interoperáveis, melhorando a troca de informações entre diferentes sistemas de saúde. Esta combinação tecnológica, principalmente, quando combinada com padrões de interoperabilidade como FHIR, é útil para construir sistemas de saúde integrados e coesos, pois atesta a compatibilidade semântica, garantindo que o significado das informações seja

uniformemente compreendido em sistemas diferentes. A aplicação do PLN em desafios de interoperabilidade na saúde impulsiona a pesquisa clínica, revela inovações e direcionamentos para futuras pesquisas, contribuindo para a extração de conhecimento de dados brutos em informações que podem ser utilizadas.

Considerando a relevância do PLN na transformação digital dos registros médicos e sua integração com padrões de interoperabilidade como FHIR, a próxima seção da dissertação se volta para a exploração detalhada do BERT, como uma das tecnologias de PLN mais avançadas atualmente.

5.3 Modelos de Aprendizado de Máquina em PLN: BERT

A Ciência da Informação sempre buscou métodos mais eficientes para organizar, recuperar e interpretar dados. Com o avanço tecnológico e a ascensão do aprendizado profundo, houve uma revolução significativa na maneira de extrair informações de grandes conjuntos de dados, especialmente, com a introdução do modelo BERT, pelo Google em 2018. Este modelo trouxe uma abordagem inovadora ao campo do PLN.

Diferentemente dos modelos anteriores, que compreendiam palavras e frases de maneira unidirecional, seja da esquerda para a direita ou vice-versa, o BERT se destacou por sua habilidade de entender o contexto bidirecional das palavras em um texto, o que significa que ele considera palavras que vêm antes e depois da palavra alvo. Essa profunda compreensão contextual foi possível graças ao treinamento do BERT, que utiliza técnicas como a previsão de palavras ocultas e a compreensão de sentenças.

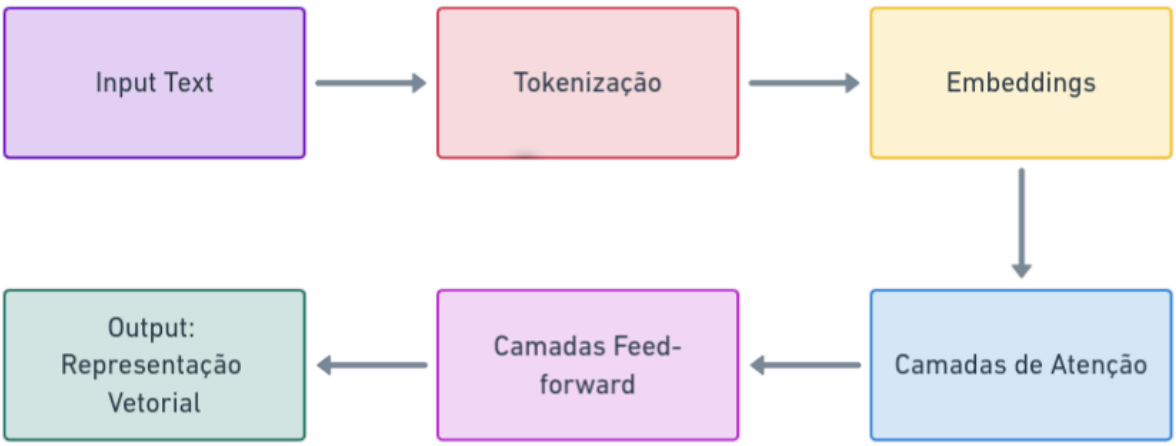
Durante o treinamento, algumas palavras são intencionalmente ocultadas, e o modelo tenta prevê-las com base no contexto das palavras circundantes. Esta capacidade avançada de compreensão contextual do BERT tem implicações diretas para a Ciência da Informação. Ao melhorar a representação semântica dos dados, ele permite uma organização e recuperação mais eficientes da informação, pois a CI se preocupa essencialmente com organização, recuperação, acessibilidade e interpretação dos dados.

Portanto, modelos como o BERT, que se alinham intimamente com estes objetivos, tornam-se ferramentas essenciais para avançar nesta área. A relevância de deste tipo de modelos na interpretação de textos complexos é reforçada pela

abordagem de Padovani (2022). A aplicação inovadora do BERT tem implicações diretas para a área da saúde, um setor que lida constantemente com grandes volumes de dados textuais complexos e, muitas vezes, não estruturados. Na saúde, a capacidade do BERT de compreender contextos bidirecionais e interpretar semanticamente os textos pode auxiliar na forma como os registros médicos eletrônicos são processados e analisados. Além disso, a habilidade do BERT em lidar com ambiguidade e nuances linguísticas, em registros médicos, contribui para uma maior precisão no diagnóstico, tratamento e pesquisa clínica, fortalecendo a interoperabilidade dos dados de saúde.

A Figura 6 retrata as etapas estabelecidas no modelo BERT.

Figura 6 – Etapas do modelo BERT.



Fonte: Adaptado de HUGGING FACE, 2024.

Na Figura 6 pode ser observado o fluxo de funcionamento do modelo BERT, com etapas que são divididas em (Hugging Face, 2024):

- *Input Text*: o texto de entrada é fornecido ao modelo. Assim como em sistemas de recuperação de informação, onde os dados são inseridos para serem processados e consultados, o texto de entrada serve como "consulta" inicial para o modelo BERT.
- *Tokenização*: o texto é dividido em *tokens* (palavras ou subpalavras) para ser processado pelo modelo. A *tokenização* pode ser comparada ao processo de indexação em bibliotecas e bancos de dados, onde informações são categorizadas e organizadas para facilitar a recuperação.

- *Embeddings*: cada *token* é convertido em um vetor de embeddings, caracterizado como uma representação numérica do *token*. Os *embeddings* são semelhantes aos metadados ou descritores, em registros de informação, fornecendo uma representação condensada e contextualizada do conteúdo original.
- Camadas de Atenção: estas camadas permitem que o modelo preste atenção a diferentes partes do texto de entrada e compreenda o contexto em que cada palavra ou *token* está inserido. Pode ser comparado ao processo de análise de relevância em sistemas de recuperação de informação, onde certas informações são priorizadas com base em sua relevância e contexto.
- Camadas *Feed-forward:* estas são camadas neurais tradicionais, que processam a informação após as camadas de atenção. Esse processo é semelhante à filtração e organização de informações em sistemas de gerenciamento de informação, onde os dados são processados e organizados em uma forma mais utilizável.
- *Output* – Representação Vetorial: o modelo BERT fornece uma representação vetorial do texto de entrada, que pode ser usada para várias tarefas. A representação vetorial pode ser vista como um resumo ou abstração do conteúdo original, semelhante a um registro bibliográfico que fornece uma visão geral do conteúdo de um documento.

O funcionamento do BERT pode ser visto como um sistema avançado de processamento e recuperação de informação, onde o texto é analisado, contextualizado e transformado em uma representação que pode ser usada para diversas tarefas de processamento de linguagem natural, refletindo muitos dos princípios fundamentais da ciência da informação. Essa compreensão serve como base para a próxima seção, que se aprofundará nos Modelos de Linguagem Grande (LLM). Os LLM representam uma evolução natural e uma expansão das capacidades ilustradas pelo BERT, oferecendo um processamento de linguagem ainda mais sofisticado e abrangente

5.4 Modelos de Linguagem Grande (LLMs)

Modelos de Linguagem Grande representam uma inovação significativa no campo do aprendizado de máquina. Esses modelos, que evoluíram rapidamente de um conceito inexistente para uma presença quase onipresente, são centrais para a compreensão e geração de linguagem natural. O desenvolvimento dos LLMs está intimamente ligado aos desafios de manipulação e análise de grandes conjuntos de dados, um aspecto essencial na Ciência da Informação. Estes modelos processam extensas quantidades de dados de texto, refletindo a necessidade de métodos eficientes de classificação, armazenamento e recuperação de informações, que são fundamentais na análise de dados (Kaddour *et al.*, 2023).

As aplicações dos LLMs estendem-se por uma ampla gama de campos, desde *chatbots* até biologia computacional, programação de computadores, trabalho criativo, direito e medicina. Em cada uma dessas áreas, os LLMs oferecem novas maneiras de lidar com informações complexas, facilitando a análise, interpretação e geração de conteúdo. Na Ciência da Informação, especificamente, os LLMs podem revolucionar a maneira como o ser humano lida com a explosão de dados na era digital, fornecendo ferramentas avançadas para análise de grandes conjuntos de dados de texto, identificação de padrões e tendências, e extração de informação. Por exemplo, na área de direito, os LLMs podem ajudar na análise e elaboração de documentos legais, refletindo a interseção entre tecnologia de processamento de linguagem natural e necessidade de gerenciamento eficiente de informações jurídicas (Kaddour *et al.*, 2023).

Especificamente na área da saúde, os LLMs estão revolucionando as práticas conhecidas e oferecendo novas possibilidades para o suporte à decisão clínica e à educação médica. Com a capacidade de processar e analisar vastas quantidades de dados médicos e científicos, os LLMs podem auxiliar profissionais de saúde na identificação de diagnósticos, tratamentos e na compreensão de complexidades médicas, com uma precisão notável, conforme destaca Singhal *et al.* (2023).

Singhal *et al.* (2023) apontam, em seu trabalho, testes com um modelo avançado de LLM, no qual atingiu um desempenho superior em 17% em relação aos testes anteriores, em questionários de estilo do exame de licença médica dos EUA. Isso ilustra o potencial dos LLMs em interpretar e responder as questões médicas complexas, abrindo caminhos para uma assistência ao paciente mais informada e

eficaz. Para além disso, os LLMs podem desempenhar um papel crucial na educação médica, fornecendo aos estudantes um recurso interativo para aprender e praticar habilidades diagnósticas e de tratamento. Essa integração de inteligência artificial na medicina pode não apenas melhorar a qualidade do atendimento ao paciente, mas também enriquecer o processo de aprendizado dos profissionais de saúde.

Portanto, os LLMs representam um avanço tecnológico significativo no campo do aprendizado de máquina e oferecem inúmeras aplicações práticas, que se entrelaçam com os princípios fundamentais da Ciência da Informação. Eles destacam a crescente interconexão entre a tecnologia de processamento de linguagem e as necessidades de gestão, análise e disseminação de informações em diversos campos (Kaddour *et al.*, 2023).

5.4.1 Modelo GPT3.5-Turbo

Após examinar a evolução geral dos Modelos de Linguagem Grande (LLMs), na subseção anterior, nesta subseção são destacadas características de um modelo específico: GPT-3.5-Turbo. Este é um modelo avançado de linguagem natural desenvolvido pela *Open*AI, representando a mais recente evolução nos LLMs (*Open*AI, 2023). Ele é caracterizado por sua capacidade de processar uma série de mensagens como entrada, ao contrário das versões anteriores que se limitavam a um único *prompt* de texto. Essa funcionalidade permite uma maior contextualização nas interações, integrando tanto o contexto fornecido quanto as respostas anteriores. O GPT-3.5-Turbo é útil por seu desempenho aprimorado e custo reduzido em comparação ao modelo GPT-3, tornando-o uma escolha eficiente e versátil para uma ampla gama de aplicações, desde *chatbots* até sistemas complexos em campos como medicina, em específico, como no caso da utilização na camada proposta.

Em termos de desempenho e custo, o GPT-3.5-Turbo se destaca por oferecer uma eficiência aprimorada, sendo 10 vezes mais econômico por *token* em comparação com seu antecessor, GPT-3 (*Open*AI, 2023). Esta melhoria não compromete a qualidade; pelo contrário, ela amplia as possibilidades de uso em uma variedade de aplicações, mantendo a adaptabilidade e a precisão, características dos LLMs. A implementação do GPT-3.5-Turbo segue um processo simples e direto, adaptável a diferentes linguagens de programação. Esta flexibilidade facilita a

integração do modelo em sistemas existentes. Na Tabela 3, apresenta-se exemplo de *prompt* de chamada para GPT-3.5-Turbo.

Tabela 3 – Exemplo de *prompt* de chamada para GPT-3.5-Turbo.

Prompt	**Resultado**	**Tempo de execução**

Dada a seguinte série de sentenças:
 	
“O paciente João da Silva tem 30 anos.”
 	
Analise as sentenças acima e identifique qual recurso FHIR mais provável em cada uma delas.
 	
Assuma o recurso mais provável para cada sentença e, em seguida, crie uma visualização JSON, sem nenhum tipo de comentário adicional.
	{
	"resourceType": "Patient",
	"name": [
	{
	"given": [
	"João"
	],
	"family": "Silva"
	}
	],
	"birthDate": "1994-01-07",
	"age": {
	"value": 30,
	"unit": "anos",
	"system": "http://unitsofmeasure.org"
	}
	}
		3 segundos

Dada a seguinte série de sentenças:
 	
“Paciente allan, 30 anos, tomou dipirona.”
 	
Analise as sentenças acima e identifique qual o recurso FHIR mais provável em cada uma delas.
 	
Assuma o recurso mais provável para cada sentença e, em seguida, crie uma visualização JSON, sem nenhum tipo de comentário adicional
	{
	"resourceType": "Patient",
	"identifier": [
	{
	"use": "usual",
	"type": {
	"coding": [
	{
	"system":
	"http://terminology.hl7.org/CodeSystem/v2-
	0203",
	"code": "MR"
	}
	]
	},
	"system": "urn:oid:1.2.36.146.595.217.0.1",
	"value": "12345"
	}
	],
	"active": true,
	"name": [
	{
	"use": "official",
	"family": "allan",
	"given": ["allan"]
	}
	],
	"gender": "male",
	"birthDate": "1991-01-01"
	},
	{
	"resourceType": "MedicationAdministration",
	"medicationReference": {
	"display": "dipirona"
	},
	"subject": {
	"reference": "Patient/example"
	}
	}
	]
		6 segundos

Fonte: Elaborado pelo autor.

A Tabela 3 exemplifica o uso do GPT3.5-Turbo via *prompt* manual e sua performance de resposta medida em segundos. É necessário destacar que o mesmo tipo de uso pode ser feito com chamadas via API.

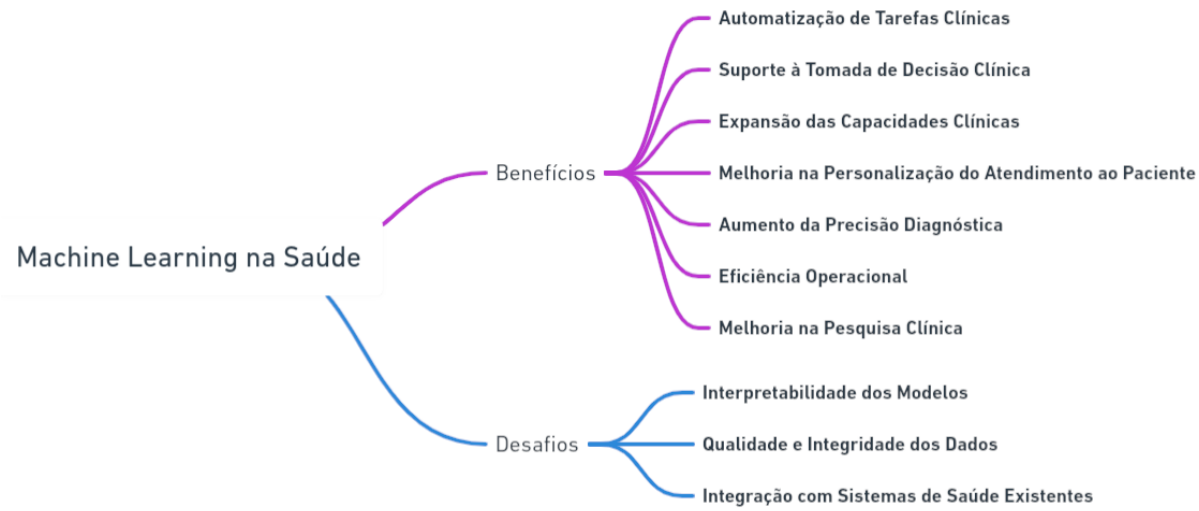
5.5 Desafios na Integração de ML e Saúde

No campo da Ciência da Informação, a combinação do M*achine Learning* com padrões de interoperabilidade, como FHIR, tem o potencial de revolucionar o modo como os dados médicos são gerenciados e utilizados. Essa combinação pode não apenas simplificar a construção de arquiteturas de sistemas robustos, mas também otimizar o armazenamento, compartilhamento e acesso à informação médica, garantindo que informações críticas estejam disponíveis quando e onde forem necessárias. No entanto, à medida em que se avança nessa integração, surgem uma série de desafios, como por exemplo, o PLN é frequentemente empregado para extrair informações de textos médicos, mas pode enfrentar dificuldades ao lidar com jargões e contextos médicos especializados, especialmente, quando se considera a diversidade e complexidade das linguagens e terminologias médicas.

Outro desafio significativo reside na necessidade de vastos conjuntos de dados rotulados para treinar modelos de ML com precisão aceitável. Em ambientes de saúde, onde privacidade e confidencialidade dos dados dos pacientes são primordiais, coletar e utilizar esses conjuntos de dados pode ser problemático. A preocupação com a ética da informação, a governança de dados e o consentimento informado dos pacientes são temas centrais na Ciência da Informação e tornam-se ainda mais pertinentes nesse contexto. Para avançar com eficácia e segurança, é necessário abordar esses desafios de forma mais ampla, considerando tanto as capacidades técnicas quanto às implicações éticas e organizacionais (Bouh; Hossain; Ahmed, 2023).

A seguir, a Figura 7 serve como uma ferramenta visual para consolidar e representar de forma estruturada os principais pontos abordados nesta seção sobre "*Machine Learning* na área da Saúde". Ao visualizar este diagrama, é possível compreender as conexões entre os diferentes tópicos, facilitando assimilação e revisão dos conceitos discutidos.

Figura 7 – *Machine Learning* na área da saúde.



Fonte: Adaptado de Ghassemi *et al.*, 2020.

A Figura 7 mostra os benefícios e desafios associados ao uso da aprendizagem de máquina na saúde, no qual, entre os benefícios, se destaca automatização de tarefas clínicas, que permite aos profissionais de saúde se afastarem de atividades repetitivas e demoradas, como a entrada de dados, aumentando a eficiência e focando mais em tarefas complexas e no atendimento ao paciente.

O *Machine Learning* também oferece suporte à tomada de decisão clínica, analisando grandes volumes de dados para fornecer diagnósticos mais precisos, prever resultados de tratamentos e identificar riscos de saúde. Além disso, também expande as capacidades clínicas ao descobrir novas correlações em dados de saúde, o que pode levar a avanços no entendimento de doenças e no desenvolvimento de novos tratamentos.

Outro benefício importante exibido é a melhoria na personalização do atendimento ao paciente, onde a análise de dados, em grande escala, permite tratamentos e recomendações mais ajustados às necessidades individuais. A precisão diagnóstica é aprimorada pelo *Machine Learning*, que analisa padrões complexos nos dados dos pacientes que podem ser difíceis de detectar manualmente. Do ponto de vista operacional, a implementação dessa tecnologia aumenta a eficiência, reduzindo custos e melhorando a gestão de recursos. No entanto, a Figura 7 também destaca os desafios associados ao uso do *Machine Learning* na área da saúde. Um dos principais obstáculos está na interpretabilidade dos modelos, considerado importante

para que os profissionais de saúde possam entender e confiar nas recomendações geradas. A qualidade e a integridade dos dados são fundamentais, pois dados imprecisos, incompletos ou enviesados podem levar a resultados inadequados ou prejudiciais.

Além disso, a integração de soluções de *Machine Learning* em sistemas de saúde existentes pode ser complexa, especialmente, devido às infraestruturas tecnológicas não interoperáveis. Portanto, embora o *Machine Learning* ofereça um potencial transformador para a saúde, é essencial abordar esses desafios para maximizar seus benefícios e minimizar riscos em potenciais.

Tendo estabelecido a relevância e o potencial do ML na área da saúde, a próxima seção se aprofunda em uma aplicação específica: a análise de dados de alergia do Hospital Sírio Libanês. A utilização de metadados de negócio se torna útil para garantir que os dados sejam interpretados corretamente e que as análises sejam precisas e forneçam contexto e significado, permitindo que os algoritmos de ML e os profissionais de saúde compreendam e usem os dados de forma mais eficaz e, dessa forma, consigam validar os modelos. Assim, é explorada como a combinação de metadados de negócio e técnicas avançadas de análise pode melhorar a qualidade e a precisão das decisões clínicas, e auxiliar na futura construção de modelos de ML.

6 DADOS DE ALERGIA DO HOSPITAL SÍRIO LIBANÊS: ANÁLISE COM METADADOS DE NEGÓCIO

Na área da saúde, a informação é uma ferramenta fundamental para a tomada de decisões clínicas, assertivas e seguras. Dentre as diversas categorias de informações médicas, os dados de alergia destacam-se como um dos elementos mais críticos, pois, conhecer as alergias de um paciente, não é apenas uma questão de precaução, mas uma necessidade essencial para evitar complicações que podem ser fatais. A reação adversa a medicamentos, alimentos ou substâncias pode variar de sintomas leves às reações anafiláticas graves, o que torna muito importante a correta identificação e registro dessas alergias.

A obtenção dos dados de alergia utilizados nesta dissertação deu-se por meio da execução de um projeto de pesquisa multidisciplinar, que inclui especialistas médicos e profissionais da computação. Este grupo de pesquisa é parte da parceria entre o Hospital Sírio Libanês e a empresa SOFYA, uma startup que tem como objetivo desenvolver soluções aplicadas na área da saúde. A SOFYA, conhecida por sua expertise em soluções de tecnologia da informação voltadas para a saúde, forneceu as ferramentas e o suporte necessários para a análise detalhada dos dados de alergia. Os profissionais do Hospital Sírio Libanês, por sua vez, contribuíram com sua profunda experiência clínica e conhecimento, validando dados e observações realizadas no domínio de alergias e suas reações adversas. Estes resultados serão publicados como parte do projeto de pesquisa que está em andamento no Instituto de Ensino e Pesquisa do Hospital Sírio Libanês.

6.1 Importância de dados de alergia

No contexto da saúde, a importância dos dados de alergia, tanto alimentares quanto medicações, é fundamental para a segurança e eficácia do atendimento ao paciente. De acordo com De la Cruz *et al.* (2018), as alergias alimentares apresentam um desafio significativo na saúde pública, afetando aproximadamente 1-3% dos adultos e 4-7% das crianças. O controle de alérgenos em alimentos é essencial, especialmente, diante da variabilidade individual na sensibilidade aos alérgenos, onde quantidades mínimas podem ser perigosas para indivíduos altamente sensíveis.

A necessidade de métodos sensíveis e confiáveis para detectar alérgenos em alimentos é importante, dada a ausência de um consenso sobre a concentração mínima de alérgenos capaz de desencadear reações alérgicas. As reações alérgicas podem variar de sintomas leves, como urticária e desconforto gastrointestinal, a reações graves, como anafilaxia, que requer intervenção médica imediata.

Reações alérgicas ocorrem quando o sistema imunológico do corpo reage de forma exagerada a uma substância estranha (alérgeno), como certos alimentos, pólen ou medicamentos. Essas reações podem variar amplamente em gravidade, desde sintomas leves, como erupções cutâneas e coceira, até condições potencialmente fatais, como anafilaxia, considerada uma reação alérgica grave e rápida, que pode afetar mais de um sistema do corpo ao mesmo tempo e requer tratamento médico imediato. É importante diferenciar entre alergia e reações alérgicas, pois a alergia é uma sensibilidade anormal a uma substância que é inofensiva para a maioria das pessoas, que, quando expostas ao alérgeno, desenvolvem uma reação alérgica como resposta. Esta distinção é fundamental para diagnóstico e tratamento adequados. Por exemplo, enquanto algumas pessoas podem ter uma alergia confirmada a um medicamento específico, outras podem experimentar reações adversas que não são mediadas pelo sistema imunológico e, portanto, não são consideradas alérgicas.

Além disso, é essencial também a capacidade de identificação de falsos positivos. Como exemplo do problema, observa-se, no estudo de Fransson *et al.* (2017), que uma parcela significativa de pacientes investigados por suspeita de alergia a medicamentos, na realidade, não apresenta alergias verdadeiras. Na pesquisa, apenas 11% dos testes resultaram em reações positivas, indicando que uma grande maioria dos pacientes suspeitos de ter alergias a medicamentos pode, de fato, tolerá-los sem riscos.

Esta observação é reforçada pelos achados de Foreman *et al.* (2020), que examinaram 43.011 registros únicos em um EHR, encontrando que 44.5% dos pacientes tinham entradas no módulo de "alergias". Foi constatado que 86.5% dessas reações foram categorizadas como alergias e 13.5% como intolerâncias. Esse estudo também revelou que 27.4% das reações documentadas utilizaram descrições em texto livre, indicando uma preferência por descrições personalizadas. Apenas 45.1% das reações consistentes com intolerância foram corretamente categorizadas como tal, ressaltando a complexidade e os desafios na documentação e categorização de reações adversas.

Portanto, é essencial a importância de abordagens detalhadas e cuidadosas no diagnóstico de alergias, seja alimentar ou aos medicamentos, na saúde. A gestão eficaz de alérgenos e a correta documentação e categorização de reações adversas são fundamentais para reduzir riscos e melhorar a qualidade do cuidado ao paciente. Estas abordagens são essenciais para o desenvolvimento de políticas de saúde pública e práticas clínicas, que visam a segurança do paciente e a prevenção de reações alérgicas adversas. A discussão sobre a precisão no diagnóstico de alergias, embasada pelos estudos de De la Cruz *et al.* (2018) e Foreman *et al.* (2020), destaca a importância da gestão eficiente de informações médicas. A próxima seção introduz a Arquitetura da Informação como uma ferramenta importante para esta finalidade, pois facilita entendimento, organização e estruturação de dados de saúde, através de mapeamento de metadados de negócio, para garantir que diagnósticos corretos sejam acessíveis em diferentes sistemas de saúde.

6.2 Aplicação de metodologia de arquitetura da informação

A aplicação de metodologia de AI tem se mostrado uma importante ferramenta para a representação, organização e acesso a informações em diversas áreas do conhecimento, incluindo a saúde, administração e sistemas legislativos. A AI permite a estruturação e modelagem da informação, de forma clara e padronizada, possibilitando o compartilhamento e a recuperação de informações entre diferentes sistemas e organizações.

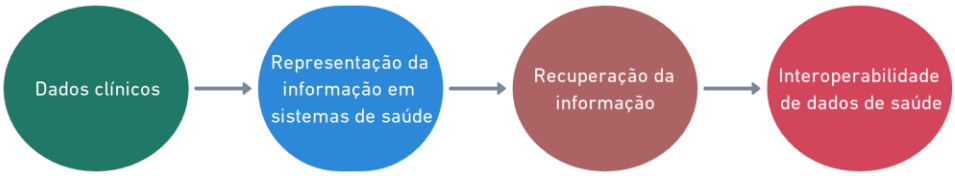
No trabalho de Brandt (2020), onde é abordada a importância do acesso à informação legislativa e à representação inadequada das informações nos sistemas da Câmara dos Deputados como um problema, apresenta-se uma metodologia de Arquitetura da Informação, proposta dividida em duas partes: elaboração da metodologia e aplicação no processo legislativo brasileiro da Câmara dos Deputados, através de mapeamento e endereçamento de metadados de negócio.

A escolha para referência deste trabalho se dá por conta de um dos artefatos apresentados, denominado como matriz de metadados, em que concentra as informações centrais necessárias para os processos de trabalho, sendo o principal entregável, pois auxilia no gerenciamento das informações, no processo de desenvolvimento de sistemas e na implementação de soluções tecnológicas. Essa

matriz reúne e dá forma ao conhecimento sobre a informação da instituição.

Brandt (2020) afirma que a representação da informação deve possibilitar que a informação seja armazenada com coerência e consistência nas bases de dados dos sistemas de informação, para permitir a gestão da informação efetiva e sua recuperação.

Figura 8 – Representação, recuperação e acesso à informação de dados clínicos.



Fonte: Adaptado de BRANDT, 2020.

A Figura 8 exemplifica o fluxo de representação e recuperação da informação. Apesar da figura original falar de outro tipo de dados, a adaptação foi útil para representação de recuperação de informação contida em prontuários eletrônicos. Baseando-se no fluxo adaptado, foi realizada uma análise sobre um conjunto de dados anonimizados fornecido pelo Hospital Sírio Libanês, no período correspondido entre 01/01/2022 e 31/07/2022, sobre avaliação de pacientes sob aspecto de alergias.

O *dataset* original possui 16 atributos, sendo composto pelos campos:

- NM_ESTABELECIMENTO: nome do estabelecimento médico. Em todos os registros listados, refere-se ao "Hospital Sírio Libanês", exceto no último registro que consta "HSL - Unidade Brasília IV".
- DS_SETOR: descreve o setor ou a unidade hospitalar em que o paciente foi atendido. Por exemplo, "Unidade Semi Intensiva (USI) - C - 06º andar", "Unidade Coronária - D - 07º andar - Ala III", etc.
- TIPO_ATENDIMENTO: tipo de atendimento que o paciente recebeu. Nos registros listados, constam "Internado" e "Pronto Atendimento".
- ADULTO_PEDIATRIA: indica se o atendimento foi para um adulto ou uma criança. Todos os registros nesta amostra são de "Adulto".
- DT_NASCIMENTO: data de nascimento do paciente.
- DT_REGISTRO: data e hora em que a informação foi registrada.

- IE_NEGA_ALERGIAS: indica se o paciente negou ter qualquer tipo de alergia. "S" indica sim (negou ter alergias) e "N" indica não (não negou ter alergias).
- IE_INTENSIDADE: refere-se à intensidade da reação alérgica, mas todos os registros, nesta amostra, têm valor "L", que pode indicar uma intensidade baixa.
- AGENTE_CAUSADOR: o que causou a alergia. Pode ser "Medicamentos", "Outro", ou especificado que o paciente "Nega Alergias" ou "Nega alergia alimentar".
- DS_PRINCIPIO4 e DS_PRINCIPIO: ambos os campos estão relacionados ao agente causador da alergia. Por exemplo, "Antibacterianos, Penicilínicos de amplo espectro", "Piridoxina", "Metoclopramida", “Paracetamol” etc.
- CLASSE_MATERIAL: classificação do material que causou a alergia. Em um dos registros está como "Contraste Radiológico".
- DS_OBSERVACAO: qualquer observação adicional sobre o paciente ou a alergia. Apenas um registro nesta amostra tem uma observação, que é "ajuste".
- DS_REACAO: descrição da reação alérgica. Por exemplo, "Erupção Cutânea", "Edema de Glote".
- FUNCAO: a função ou profissão da pessoa que registrou a informação. Pode ser "Enfermeiro Pl", "NUTRICIONISTA CLINICO", "MEDICO".
- TIPO_ATENDIMENTO: esta coluna está repetida, já que já foi mencionada anteriormente na lista.

A seguir, a Tabela 4 apresenta uma amostra do *dataset* original:

Tabela 4 – Amostra original dos dados de alergia do HSL.

#	1	2	3	4	5	6	7
NM ESTABELECIMENTO	HSL	HSL	HSL	HSL	HSL	HSL	HSL
DS SETOR	Unidade Semi Intensiva (USI) - C - 06º andar	Unidade Coronária - D - 07º andar - Ala III	Unidade Internação - D - 17º andar - Ala II	Unidade Internação – D - 18º andar - Ala II	D - UAIC - Ala VI	D - UAIC - Ala VI	D - UAIC - Ala VI
TIPO ATENDIMENTO	Internado	Internado	Internado	Internado	Internado	Internado	Internado
ADULTO PEDIATRIA	Adulto	Adulto	Adulto	Adulto	Adulto	Adulto	Adulto
DT NASCIMENTO	11/27/1981	1/19/1969	7/24/1936	7/17/1929	9/28/1961	9/28/1961	9/28/1961
DT REGISTRO	1/1/21 11:54	1/1/21 11:56	1/1/21 12:11	1/1/21 12:12	1/1/21 12:13	1/1/21 12:13	1/1/21 12:13
IE NEGA	N	N	S	S	N	N	N
IE INTENSIDADE	L	L	L	L	L	L	L
AGENTE CAUSADOR	Nega Alergias	Medicamentos			Outro	Medicamentos	Medicamentos
DS PRINCIPIO4	-	Antibacterianos, Penicilinícos de amplo espectro	Nega alergia alimentar	Nega alergia alimentar	piridoxina	-	-
DS PRINCIPIO	-	-	-	-	-	Metoclopramida	Paracetamol
CLASSE MATERIAL	-	Antibacterianos, Penicilinícos de amplo espectro	-	-	-	-	-
DS OBSERVAÇÃO	-	-	-	Ajuste	-	-	-
DS REACAO	-	Erupção Cutânea	-	-	Erupção Cutânea	Edema de Glote	Erupção Cutânea
FUNCAO	Enfermeiro Pl	Enfermeiro Pl	Nutricionista Clínico	Nutricionista Clínico	Médico	Médico	Médico
TIPO ATENDIMENTO	Internado	Internado	Internado	Internado	Internado	Internado	Internado

Fonte: Elaborado pelo autor.

Após análise inicial dos dados e entrevista com os profissionais, que forneceram o material, foi verificado que somente 10 atributos eram relevantes para a análise de metadados de negócio. Dessa forma, o número de colunas foi reduzido e renomeado para os seguintes:

- Setor;
- Tipo de atendimento;
- Faixa etária;
- Data nascimento;
- Data de registro;
- Intensidade;
- Agente causador;
- Princípio;

- Reação;
- Função.

Após a redução, foram realizadas análises dos atributos contidos no *dataset*, a fim de identificar os conceitos e suas relações, gerando um mapa conceitual, que pode ser observado na Figura 9.

Figura 9 – Mapa conceitual dos atributos do *dataset* de alergia do HSL.

Fonte: Elaborado pelo autor.

Mediante aos atributos e relações dispostas, foi realizada uma pesquisa no repositório do padrão FHIR para encontrar qual recurso melhor se adequaria aos dados propostos. O recurso *AllergyIntolerance*, do padrão FHIR, foi concebida com o propósito de representar informações relativas às alergias ou intolerâncias, que um indivíduo pudesse ter a um determinado estímulo, como um medicamento, alimento, ou qualquer outra substância (HL7 FHIR, 2021).

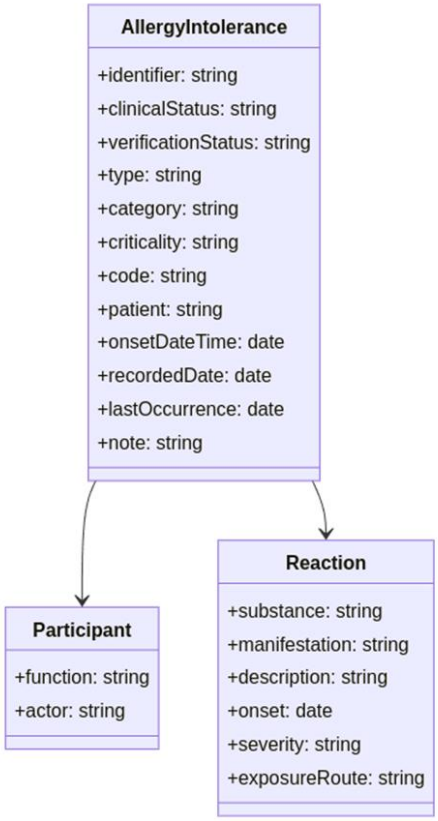
Ao analisar o conjunto de dados do HSL, foi observado campos como AGENTE_CAUSADOR, DS_PRINCIPIO4, DS_PRINCIPIO, e CLASSE_MATERIAL. Esses campos indicaram claramente a substância ou agente causador da alergia. Foi identificado que tais campos mapeavam-se diretamente e de maneira intuitiva com o atributo *code*, do recurso *AllergyIntolerance*, que especifica a substância alergênica.

A descrição da reação alérgica, representada no campo DS_REACAO, do conjunto de dados, teve sua correspondência no recurso *AllergyIntolerance*, no

atributo *reaction*. Esse atributo permitia que fossem detalhadas não apenas a manifestação da alergia, mas também sua gravidade, seu início e outros aspectos relevantes. O campo IE_INTENSIDADE, que indica a intensidade da reação alérgica, foi mapeado para o atributo *criticality*, do recurso FHIR, proporcionando uma visão sobre a severidade potencial da reação adversa. Por fim, foi destacado o aspecto temporal das alergias. No conjunto de dados, DT_REGISTRO especifica quando a informação foi registrada, correlacionando-se com o atributo *recordedDate* do recurso *AllergyIntolerance*.

A capacidade de rastrear quando uma alergia foi identificada e registrada mostrou-se crucial para gestão clínica e para entender a evolução do paciente ao longo do tempo. Ao considerar todas essas correspondências diretas e a natureza intrínseca do recurso *AllergyIntolerance* para representar alergias e intolerâncias, tornou-se evidente que esta era o recurso FHIR mais adequado para assegurar a interoperabilidade dos dados analisados.

Figura 10 – Representação do recurso FHIR *AllergyIntolerance*.



Fonte: Adaptado de http://hl7.org/FHIR/allergyintolerance.html.

A Figura 10 apresenta uma representação abstraída do recurso FHIR *AllergyIntolerance*, um recurso central na modelagem de informações relativas a alergias ou intolerâncias em um contexto clínico. Esta representação é de suma importância para compreender a estrutura e os relacionamentos inerentes ao recurso dentro do padrão FHIR, fundamental para garantir a interoperabilidade de dados de saúde. No centro da figura, identifica-se o recurso *AllergyIntolerance*, que é circundada por seus atributos-chave. Estes são:

- *Identifier*: um identificador único para a intolerância ou alergia.
- *ClinicalStatus*: descreve o estado clínico atual do registro.
- *VerificationStatus*: indica o nível de verificação da alergia ou intolerância.
- *Type*: o tipo de alergia ou intolerância.
- *Category*: a categoria da substância alergênica.
- *Criticality*: indica o potencial de risco associado à reação.
- *Code*: especifica a substância alergênica.
- *Patient*: associa o registro à identidade do paciente em questão.
- *Encounter*: refere-se à interação clínica durante a qual a alergia ou intolerância foi registrada.
- *Onset*: informa o início da alergia ou intolerância.
- *RecordedDate*: data em que a alergia ou intolerância foi registrada.
- *Lastocurrence*: a última ocorrência da reação.
- *Note*: quaisquer observações ou anotações adicionais sobre a intolerância ou alergia.

Associado ao *AllergyIntolerance*, há um recurso chamado *Participant*, que detalha quem registrou ou observou a alergia ou intolerância. Ele possui dois atributos principais:

- *Function*: especifica a função do participante, por exemplo, médico, enfermeiro ou outro profissional de saúde.
- *Actor*: identifica o indivíduo ou entidade específica que desempenha essa função.

Além disso, há uma associação com recurso *Reaction*, que descreve a reação específica que o paciente teve à substância alergênica. Os atributos deste recurso incluem:

- *Substance*: substância específica que causou a reação.
- *Manifestation*: manifestação clínica da reação.
- *Description*: descrição detalhada da reação.
- *Onset*: início da reação.
- *Severity*: gravidade da reação.
- *ExposureRoute*: meio pelo qual o paciente foi exposto à substância.
- *Note*: quaisquer observações ou anotações adicionais sobre a reação.

Essa representação UML proporciona uma visão clara e estruturada do recurso FHIR *AllergyIntolerance*, e de como as informações sobre alergias e intolerâncias são estruturadas e relacionadas dentro do padrão FHIR.

Uma vez encontrada o recurso FHIR alvo, que permite a interoperabilidade dos dados do *dataset,* a análise foi continuada a fim de ter uma visão mais ampla do ponto de vista processual, a fim de identificar não só o contexto de produção de cada atributo, mas também do ponto de vista de negócios.

No contexto empresarial, cada campo de informação pode ter implicações significativas para tomada de decisão, gestão de recursos e estratégias de operação. Sem uma compreensão clara do propósito e do contexto de negócios associado a cada campo, as análises derivadas desses dados podem ser mal interpretadas ou mal aplicadas. Além disso, para alcançar uma interoperabilidade eficaz entre sistemas e plataformas, é fundamental que as partes interessadas compartilhem uma compreensão unificada, não apenas do que os dados representam semanticamente, mas também de seu valor e função no contexto de negócios.

Uma compreensão profunda e clara dos campos de dados facilita a integração de informações entre sistemas, promove a clareza na comunicação e garante que as informações sejam utilizadas de maneira otimizada, para gerar valor real para as organizações.

Neste sentido, Brandt (2020) apresenta um artefato chamado Matriz de Metadados, onde propõe que é o principal recurso para uma boa Arquitetura da Informação. Ele é composto pela extração e identificação dos seguintes elementos:

- Nome: refere-se ao título ou designação específica do metadado ou campo de informação. Basicamente, é o rótulo que identifica um dado específico.
- Contexto: descreve a situação ou cenário em que o dado é utilizado ou o

propósito para o qual foi criado. Pode incluir a função ou significado mais amplo do dado, dentro de um sistema ou processo.

- Definição: fornece uma descrição clara e concisa do que exatamente o dado representa ou significa. Ajuda a garantir que todos os usuários ou interessados compreendam o dado da mesma maneira.
- Gestor: identifica a pessoa ou entidade responsável pela administração e manutenção geral do dado. Esta é a autoridade que garante a integridade e precisão do dado.
- Gestor do dado: é a pessoa ou grupo que tem a responsabilidade primária pelo dado, incluindo sua coleta, atualização e precisão. Eles podem, ou não, ser os criadores originais do dado, mas são responsáveis por sua manutenção contínua.
- Forma de Acesso: descreve como os usuários ou sistemas podem acessar ou recuperar o dado. Isso pode incluir métodos, plataformas ou ferramentas específicas.
- Entrada Padronizada: especifica se o dado deve ser inserido de uma maneira particular ou padronizada, garantindo consistência em todo o sistema.
- Regra de Formato: define os padrões ou diretrizes sobre como o dado deve ser apresentado ou formatado. Isso pode se referir a coisas como tamanho, tipo ou estrutura do dado.
- Alimentação: descreve como o dado é coletado, inserido ou alimentado no sistema. Pode referir-se aos métodos manuais, automáticos ou uma combinação de ambos.
- Dados Abertos: indica se o dado é de acesso público e pode ser acessado, usado e compartilhado por qualquer pessoa. Os dados abertos são frequentemente usados para promover a transparência e a colaboração.
- Restrição de Acesso: se existir, esta coluna indica quaisquer limitações ou restrições sobre quem pode acessar o dado, quando e sob quais condições. As restrições podem ser devido às questões de privacidade, segurança ou outras razões regulatórias.

Considerando a Matriz de Metadados proposta e o contexto do trabalho, ajustes foram realizados, em que duas colunas adicionais foram acrescentadas: a primeira,

para identificação do "Vocabulário de saúde destino", considerando que os recursos do padrão FHIR podem ser compatíveis com diferentes terminologias; e a segunda, para a "Nomenclatura do conceito FHIR", onde a informação do atributo correspondente é armazenada. Com esses aprimoramentos na matriz, garantiu-se uma representação ainda mais robusta e alinhada às necessidades de interoperabilidade. Após essa formatação, o produto final foi uma Matriz de Metadados, disponibilizada na Tabela 5 a seguir:

Tabela 5 – Metadados de negócio gerados com interoperabilidade com padrão FHIR.

ID	1	2	3	4	5	6	7	8	9	10
Nome	Setor	Tipo Atendimento	Adulto Pediatria	Data Nascimento	Data Registro	Intensidade	Agente Causador	Princípio	Reação	Função
Contexto	Identificar a equipe de atendimento	Identificar a modalidade de atendimento	faixa etária do paciente	Identificar idade do paciente	Identificar o período do atendimento	Indica a severidade da ocorrência do sintoma	Identificação do mecanismo fisiológico para um Risco de Reação	Identificação do elemento causador das manifestações	Identificar as reações ou manifestações es que evidenciam a presença de alergia	Identificar quem ou o que participou das atividades relacionadas à alergia ou intolerância e como se envolveu
Definição	Diferentes equipes que realizaram os procedimentos de avaliação com o objetivo de identificação de alergias.	Diferentes modalidades para identificar as acomodações e agendamento do contato assistencial	Indica o tipo de cuidado assistencial necessário mediante a idade	Dia, mês e ano do nascimento do paciente	Dia, mês e ano do nascimento do contato assistencial	Diferentes níveis de manifestações alérgicas	Diferentes origens causadoras do sintoma/ reação alérgica	Diferentes produtos, alimentos responsáveis pelo desencadeamento	Diferentes eventos de reação adversam ligados à exposição à substância	Diferentes funções dos profissionais de saúde
Gestor	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial
Gestor de dados	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial	Coordenação ambulatorial
Forma de acesso	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL
Entrada padronizada	Semi Intensiva (USI) - D - 11º andar Ala I | Unidade Semi Intensiva (USI) - D - 11º andar Ala II | Unidade Semi Intensiva (USI) - D - 11º andar Ala III | C - UTI Adulto - Ala I | C	Pronto Atendimento | Internado	Adulto Pediatria	Não se aplica	Não se aplica	I|D|L|M	*Food* | *medication* | *environment* | *biologic*	Não se aplica	Agitação psicomotora | Angioedema | Broncoespasmo | Choque anafilático | Coriza| Dermatite | Edama | Edema de glote | Erupção cutânea | Intolerância| Parada cardíaca | Prurido| Tosse	Auxiliar Enfermagem | Assistente atendimento | Enfermeiro | Farmacêutico | Médico Médico Residente | Nutricionista Clínico | Preceptor | Técnico Enfermagem
Regra formato	Textual	Textual	Textual	Data	Data	Data	Textual	Textual	Textual	Textual
Alimentação inicial	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL	ERP HSL
Dados abertos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
Restrição de acesso	Privado	Privado	Privado	Privado	Privado	Privado	Privado	Privado	Privado	Privado
Conceito FHIR	Não se aplica	Não se aplica	Não se aplica	*Patient:birthDate*	*AllergyIntolerance :recordedDate*	*llergyIntoleranc e:Manifestation:s everity*	*AllergyIntolerance:cat egory*	*AllergyIntolerance:sub stance*	*AllergyIntolerance:reactio n:substance:codi ng:code*	*AllergyIntolerance:Par ticipant:function*
Vocabulário	-	-	-	-	-	*fhir/ValueSet/reac tion- event-s everity*	*fhir/ValueSet/allergyin tolerance -category*	SNOMED	SNOMED	*fhir/ValueSet/partici patio n-role-ty pe*

Fonte: Elaborado pelo autor.

Após a conclusão da análise dos dados de alergia fornecidos pelo Hospital Sírio Libanês, evidenciou-se a importância dos metadados de negócio no contexto dos sistemas de saúde. Essa análise demonstrou que, sem uma estruturação e interoperabilidade adequadas dos dados, a qualidade e aplicabilidade da informação podem ser comprometidas. O papel dos metadados no ambiente de negócios foi compreendido, ressaltando sua contribuição essencial para assegurar a informação correta.

Com base nesse trabalho já realizado, o próximo passo do estudo, detalhado na seção subsequente, é destinado à criação de um modelo informacional para a interoperabilidade de dados de saúde. Nessa fase, o *Machine Learning* é empregado para mapear dados conforme o padrão FHIR, permitindo uma abordagem mais avançada e precisa para integração dos sistemas de saúde.

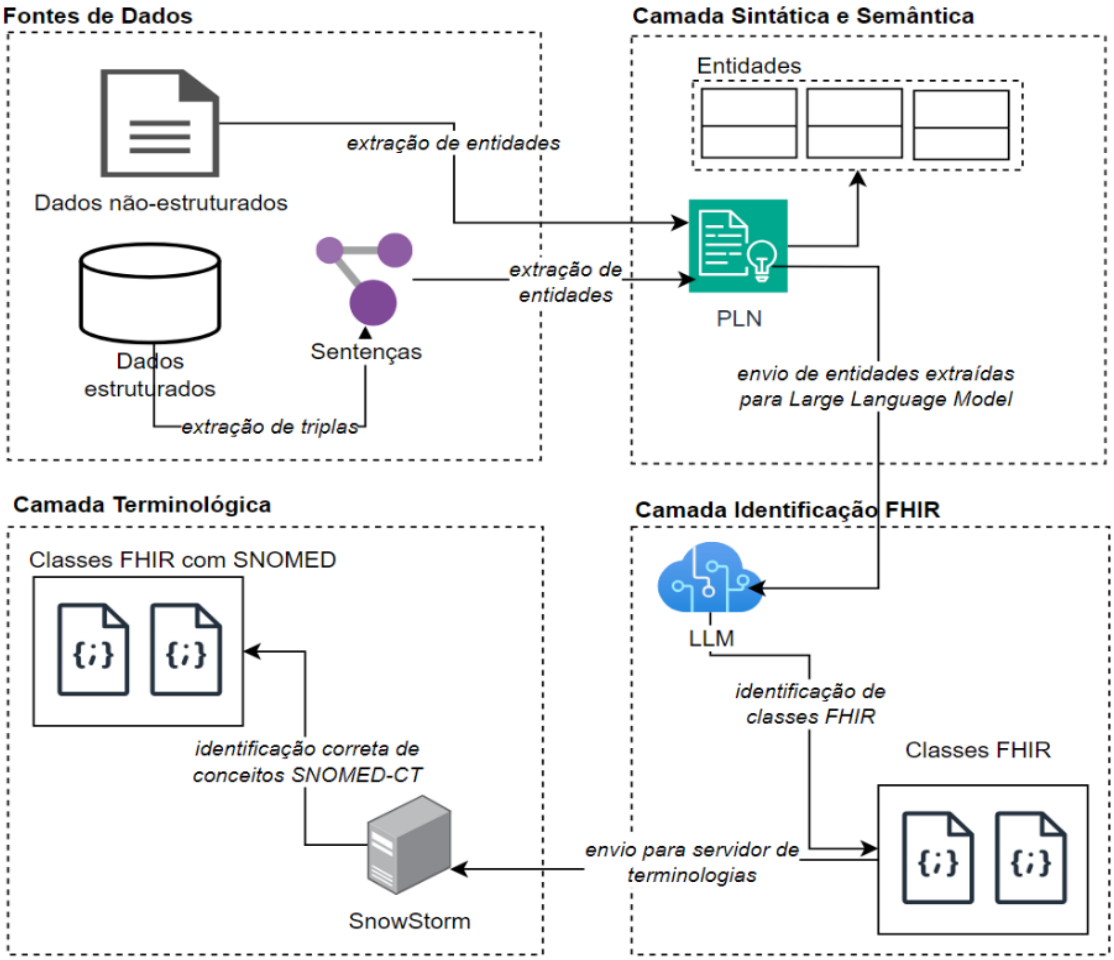
7 FHIR-FLOW: MODELO DE MAPEAMENTO DE DADOS DE SAÚDE PARA FHIR

Esta seção introduz um modelo de interoperabilidade de dados de saúde, projetado especificamente para mapeamento de dados no padrão FHIR, denominado "FHIR-FLOW”. Este modelo foi desenvolvido para abordar e solucionar os desafios de interoperabilidade de dados, que são críticos no campo da saúde. O principal objetivo é facilitar a troca eficiente e precisa de informações de saúde entre diferentes plataformas e sistemas, melhorando assim a qualidade do atendimento ao.

O FHIR-FLOW se tem como característica a capacidade de processar e interpretar diferentes dados de saúde, tanto estruturados quanto não estruturados, e convertê-los em um formato padronizado e interoperável. Utilizando tecnologias avançadas de Processamento de Linguagem Natural e Modelos de Linguagem Grande, como o GPT-4, ele é capaz de analisar, estruturar e mapear dados de saúde com alta precisão.

Esta seção detalha as três etapas principais do modelo: análise sintática e semântica, identificação FHIR e camada terminológica; e explora suas aplicações práticas, desafios e potencial impacto no campo da saúde.

Figura 11 – Fluxo geral de transformações dos dados e etapas do FHIR-FLOW.



Fonte: Elaborado pelo autor.

A Figura 11 ilustra o funcionamento do FHIR-FLOW, um modelo desenhado para atuar diretamente sobre os dados já persistidos nas unidades de saúde, abordando-os em sua forma original, sejam eles estruturados ou não. Este modelo inovador serve como uma ponte eficaz, facilitando a conversão de dados brutos e heterogêneos em informações padronizadas e interoperáveis, sem a necessidade de processos de transformação prévios.

Ao iniciar sua operação, o FHIR-FLOW aplica uma análise sintática para normalizar os dados, preparando-os para um processamento mais avançado. Através da utilização de Modelos de Linguagem Grande e técnicas avançadas de reconhecimento de entidades (NER), o modelo tem a capacidade de compreender, interpretar e gerar textos mantendo relevância e coerência, identificando entidades-chave e suas relações dentro dos dados. Progressivamente, ele incorpora camadas de identificação semântica e terminológica, enfocando nos recursos FHIR e na correta

identificação de terminologias como SNOMED CT. Este procedimento assegura que, independentemente do formato original dos dados nas fontes de saúde, o FHIR-FLOW garante consistência, alinhamento e interoperabilidade das informações processadas.

Na primeira etapa, na análise sintática, há um processo para tratar dados estruturados e não estruturados, como normalizar e estruturar dados em forma de sentenças e triplas, preparando-os para formação de textos que, subsequentemente, serão processados com técnicas de PLN. Para este processo, são empregados LLM, que são capazes de compreender e gerar textos com alto grau de coerência e relevância. O resultado desta etapa são entidades nomeadas e suas relações, extraídas e identificadas através de triplas.

Na segunda etapa, na análise semântica, é introduzida uma camada de identificação de recursos FHIR, essencial para interpretação e mapeamento corretos dos dados, garantindo que as informações sejam interpretadas e representadas de forma unificada, respeitando os padrões e normativas do FHIR. Na etapa terminológica, o FHIR-FLOW foca na classificação e padronização de terminologias, ao ressaltar a SNOMED CT, através do uso de um servidor de terminologias especializado em SNOMED CT. Esta etapa é essencial para garantir que as terminologias sejam uniformemente compreendidas e aplicadas, independentemente das variações terminológicas entre diferentes sistemas, promovendo interoperabilidade.

O resultado final do modelo proposto são recursos FHIR gerados a partir de dados, sejam eles estruturados ou não estruturados, permitindo uma comunicação integrada e harmonizada entre sistemas de informação em saúde distintos.

7.1 Camada Sintática e Semântica

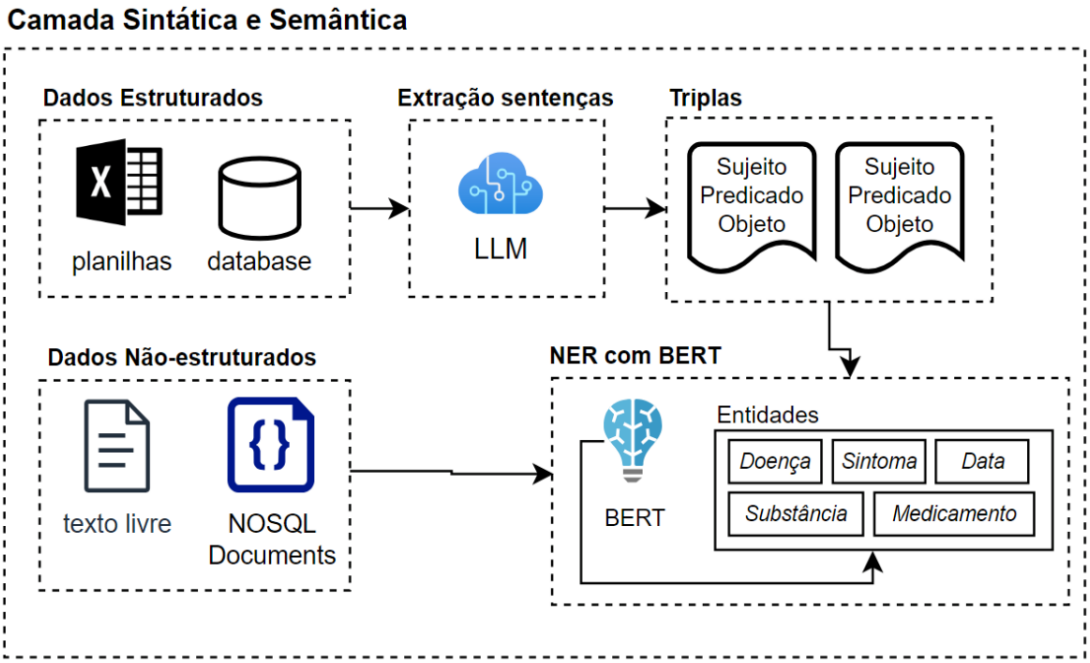
Nesta subseção é apresentada a primeira etapa do FHIR-FLOW. Esta camada é a fundação sobre a qual o modelo constrói sua capacidade de interpretar e processar dados de saúde. A ênfase é colocada na análise e na estruturação dos registros de saúde, tanto estruturados quanto não estruturados, para facilitar sua interpretação e integração posterior. Essencialmente, a camada sintática e semântica serve como ponto de entrada para os dados no sistema, onde são submetidos ao processo de normalização e análise gramatical.

O objetivo principal desta camada é transformar os registros eletrônicos de saúde em formatos que são coerentes para a interpretação humana e compatíveis

com processamentos automáticos posteriores. Esse processo envolve a desagregação de dados estruturados e não estruturados em componentes mais simples, como sentenças e frases, e a aplicação de técnicas avançadas de PLN, para extrair significado e contexto dos dados não estruturados. Nesse sentido, o modelo BERT é utilizado para identificar e classificar entidades e relações dentro do texto, preparando o terreno para análises mais complexas nas etapas subsequentes do modelo.

Ao final desta fase, o modelo gera um conjunto de triplas, que captura as relações entre entidades identificadas, que, por sua vez, são requisitos para o mapeamento de dados para recursos FHIR, pois representam a estrutura básica das informações que serão utilizadas nas próximas etapas do modelo.

Figura 12 – Etapas da extração sintática e semântica.



Fonte: Elaborado pelo autor.

A Figura 12 detalha o fluxo de extração sintática/semântica com as seguintes etapas:

- Entrada de EHR: os registros eletrônicos de saúde são inseridos no sistema.
- Dados Estruturados: estes passam por uma etapa de extração de sentenças para converter os dados estruturados em sentenças individuais.
- Dados Não Estruturados: estes são direcionados diretamente para o reconhecimento de entidades usando o modelo BERT.
- Extração de Sentenças a partir de Dados Estruturados: esta etapa converte dados estruturados em sentenças individuais.
- Reconhecimento de Entidades com BERT: tanto as sentenças extraídas dos dados estruturados, quanto os dados não estruturados, são alimentados no modelo BERT para o reconhecimento de entidades.

O resultado desta etapa é um conjunto de triplas identificando relações entre entidades, como mostrado no exemplo da Tabela 6 abaixo, com dados sintéticos gerados a partir de LLM:

Tabela 6 – Exemplo de extração de triplas a partir de fontes de dados.

<table>
<tr><th colspan="5">Dado Original</th><th>Triplas</th></tr>
<tr><td colspan="5">João, 29 anos, relatou desconforto e prurido após consumir um sanduíche com pasta de amendoim. Ele não tinha histórico de consumo de amendoim ou outras alergias alimentares conhecidas e mantinha um estado de saúde geralmente bom. No exame, João apresentava sinais vitais estáveis, sinais de urticária e inchaço labial.

Diante dos sintomas, suspeita-se de reação alérgica ao amendoim. Exames alergológicos, incluindo teste cutâneo e dosagem de IgE específica para amendoim, foram solicitados para confirmação. João foi orientado a evitar amendoim e recebeu antialérgico para aliviar os sintomas, com instruções sobre adrenalina auto injetável para possíveis casos de anafilaxia futura.</td><td>Paciente, tem idade de, 29 anos.
Paciente, relatou, desconforto.
Paciente, relatou, prurido.
Paciente, consumiu, sanduíche com pasta de amendoim.
Paciente, não tinha, histórico de consumo de amendoim.
Paciente, não tinha, histórico de outras alergias alimentares conhecidas.
Paciente, mantinha, estado de saúde geralmente bom.
Paciente, apresentava no exame, sinais vitais estáveis.
Paciente, apresentava no exame, sinais de urticária.
Paciente, apresentava no exame, inchaço labial.
Paciente, foi orientado a, evitar amendoim.
Paciente, recebeu, antialérgico para aliviar os sintomas.
Paciente, recebeu, instruções sobre adrenalina auto injetável para possíveis casos de anafilaxia futura.
Exames alergológicos, foram solicitados para, confirmação de reação alérgica ao amendoim.
Teste cutâneo e dosagem de IgE específica para amendoim, foram incluídos em, exames alergológicos.</td></tr>
<tr><td>ID</td><td>Data</td><td>CID</td><td>Substância</td><td>Sintoma</td><td rowspan="5">Paciente, apresentou Desconforto devido a, Pasta de amendoim.
Paciente, apresentou Prurido devido a, Pasta de amendoim.
Paciente, apresentou Sinais de urticária devido a, Pasta de amendoim.
Paciente, apresentou Inchaço labial devido a, Pasta de amendoim.</td></tr>
<tr><td>1335</td><td>2021-09-10</td><td>F1001</td><td>Pasta de amendoim</td><td>Desconforto</td></tr>
<tr><td>1335</td><td>2021-09-10</td><td>F1001</td><td>Pasta de amendoim</td><td>Prurido</td></tr>
<tr><td>1335</td><td>2021-09-10</td><td>F1001</td><td>Pasta de amendoim</td><td>Sinais de urticária</td></tr>
<tr><td>1335</td><td>2021-09-10</td><td>F1001</td><td>Pata de amendoim</td><td>Inchaço labial</td></tr>
</table>

A análise sintática foca na estrutura gramatical dos textos, identificando relações entre palavras e frases para determinar seu significado. Em registros clínicos, por exemplo, a análise sintática pode diferenciar entre "febre" como um sintoma e "febre" como parte de uma frase mais ampla, como "sem febre por três dias". Esta

capacidade de discernir contextos e relações é comparável à catalogação em Ciência da Informação, onde itens são organizados e classificados com base em suas características e relações inerentes. Assim como um bibliotecário organiza livros em prateleiras para facilitar a recuperação de informações, a análise sintática organiza e estrutura dados de texto, para torná-los mais acessíveis e compreensíveis. Esta analogia destaca a interseção entre os princípios da CI e as técnicas de PLN, demonstrando a importância de abordagens estruturadas na interpretação de dados não estruturados.

A análise semântica se dedica a decifrar os significados de palavras e sentenças, interpretar contextos e desvendar ambiguidades linguísticas, promovendo uma compreensão mais profunda e contextualizada dos textos. Este processo é essencial para extração, interpretação e organização eficiente de informações, particularmente, em campos onde a precisão da informação é crucial, como na saúde.

Dentro dela, a identificação de entidades nomeadas, ou *Named Entity Recognition* (NER), é uma técnica importante que identifica e classifica entidades em textos, tais como pessoas, organizações, locais e datas. Ela é essencial para extração de conhecimento em NLP, com aplicações em várias áreas, como extração de informação, busca de informações, tradução automática, entre outras (Liu *et al*., 2022).

Na Ciência da Informação, o NER é útil para organização e recuperação de dados, que permite a indexação semântica de documentos, facilitando a recuperação de informações, baseada em conceitos e promovendo a descoberta de conhecimento. Esta técnica potencializa a construção de metadados semânticos e aprimora os sistemas de informação, tornando-os mais intuitivos e eficientes.

No contexto da saúde, a identificação de entidades nomeadas é, especialmente, relevante, pois pode impactar diretamente na qualidade e eficácia do atendimento ao paciente, pois, como destaca Yang *et al.* (2022), os dados de texto médico registram dados clínicos detalhados e o reconhecimento de entidades nomeadas é a base do processamento de informações textuais, no qual é considerada uma parte importante da extração de informações.

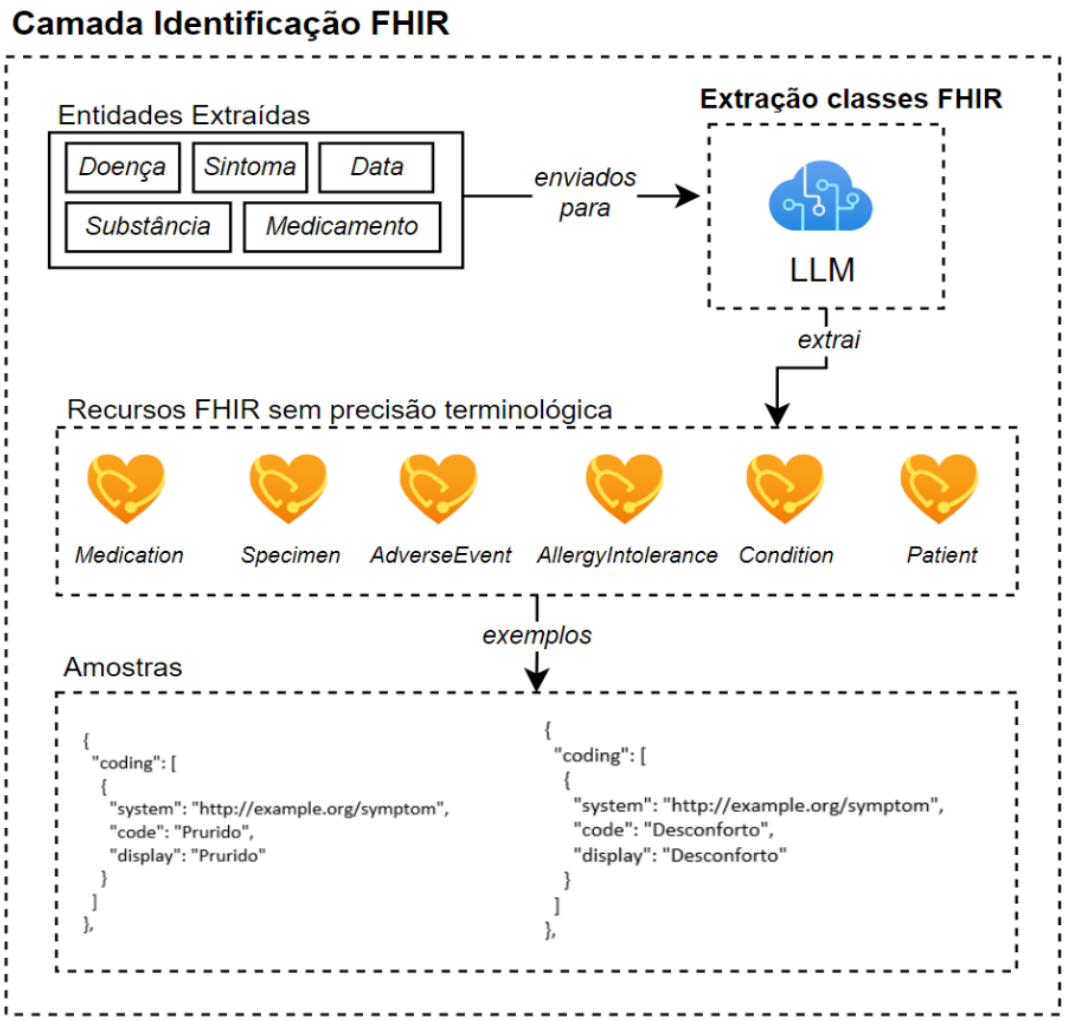
A extração precisa de informações, como nomes de medicamentos, diagnósticos, procedimentos médicos e condições de saúde, de textos não estruturados, pode facilitar análises, interpretações e identificação de contextos clínicos. Este processo é essencial para a atribuição correta de textos clínicos em recursos FHIR, que são divididas por contextos e entidades específicas.

Assim, se um relatório médico menciona que um paciente foi diagnosticado com hipertensão e prescrito um medicamento específico, o sistema de NER pode reconhecer "hipertensão" como uma *Condition* (Condição) e o nome do medicamento como uma *Medication* (Medicação), atribuindo, consequentemente, estas informações aos recursos FHIR correspondentes.

7.2 Camada Identificação FHIR

Nesta subseção é explorada a camada de identificação FHIR, que foca na categorização e mapeamento apropriados dos dados de saúde para os recursos do padrão FHIR. Os recursos FHIR são essenciais para organização de dados em categorias específicas, como Recursos, Extensões, Perfis e Operações. Nesta etapa do FHIR-FLOW, utilizou-se o GPT-4 da *Open*AI, um modelo de linguagem grande, para analisar as triplas de dados compostas por sujeito, predicado e objeto extraídas na etapa anterior. Estas triplas, que representam as relações semânticas entre diferentes entidades nos dados de saúde, são submetidas ao GPT-4 para a identificação FHIR. O uso do GPT-4 é fundamental nesta fase, pois seu conhecimento semântico e contextual avançado é aplicado para predizer, com alta precisão, os recursos FHIR mais adequados para cada tripla, porém sem a responsabilidade de precisão terminológica.

Figura 13 – Camada de identificação FHIR.



Fonte: Elaborado pelo autor.

Conforme ilustra a Figura 13, nesta etapa, foi utilizado um modelo de LLM, o GPT-4 da *Open*AI (*Open*AI, 2023), para analisar e mapear as triplas aos respectivos recursos FHIR. As triplas, contendo sujeito, predicado e objeto, são extraídas da camada anterior e representarão as relações semânticas entre diferentes entidades.

O processo foi desenvolvido da seguinte maneira: inicialmente, as triplas serão extraídas e submetidas a um pré-processamento, no qual serão limpas e normalizadas, removendo ruídos e convertendo as entidades para um formato compatível com o modelo de linguagem. Posteriormente, o modelo de linguagem é alimentado com as triplas pré-processadas, gerando predições sobre os possíveis recursos FHIR correspondentes. O conhecimento semântico e contextual do modelo de linguagem permite a associação das triplas aos recursos FHIR mais prováveis.

A seguir, a Tabela 7 exemplifica uma entrada e saída desta etapa de identificação FHIR:

Tabela 7 – Exemplo de identificação de recursos FHIR através de triplas.

Triplas	Recursos FHIR

Paciente, apresentou Desconforto devido a, Pasta de amendoim.
Paciente, apresentou Prurido devido a, Pasta de amendoim.
Paciente, apresentou Sinais de urticária devido a, Pasta de amendoim.
Paciente, apresentou Inchaço labial devido a, Pasta de amendoim.
	{
	"resourceType": "AllergyIntolerance",
	"patient": {
	"reference": "Patient/example",
	"display": "Paciente"
	},
	"substance": {
	"coding": [
	{
	"system": "http://example.org/substance",
	"code": "PastaDeAmendoim",
	"display": "Pasta de Amendoim"
	}
	]
	},
	"reaction": [
	{
	"manifestation": [
	{
	"coding": [
	{
	"system": "http://example.org/symptom",
	"code": "Desconforto",
	"display": "Desconforto"
	}
	]
	},
	{
	"coding": [
	{
	"system": "http://example.org/symptom",
	"code": "Prurido",
	"display": "Prurido"
	}
	]
	},
	{
	"coding": [
	{
	"system": "http://example.org/symptom",
	"code": "InchacoLabial",
	"display": "Inchaço Labial"
	}
	]
	}
	]
	}
	]
	}

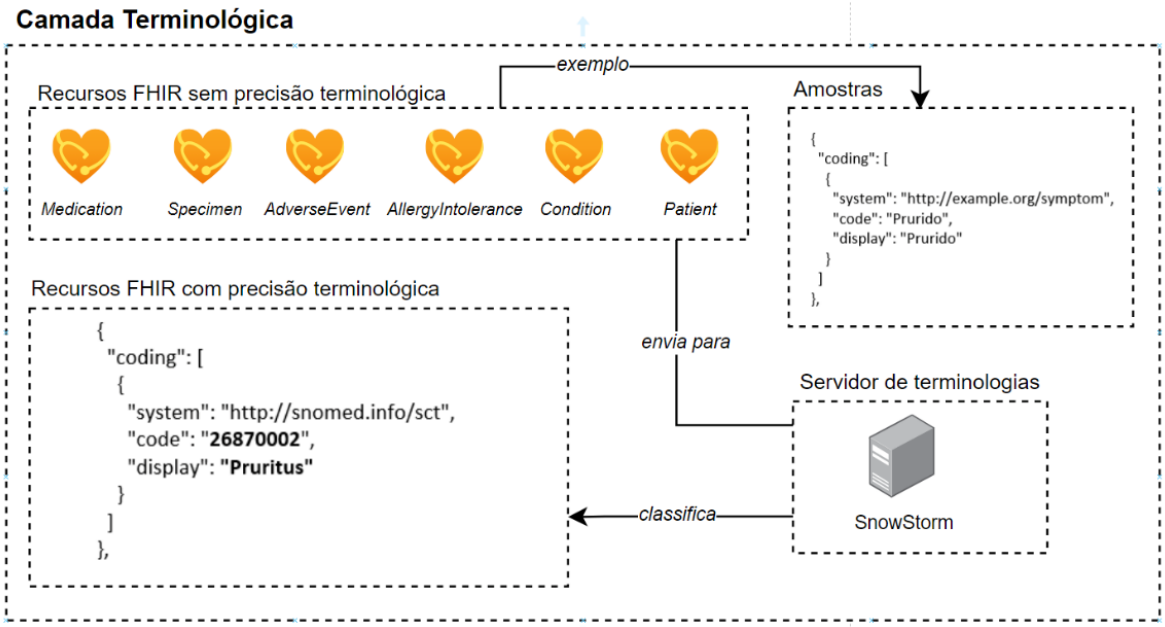
Fonte: Elaborado pelo autor.

7.3 Camada Terminológica

Esta subseção explica o processo para identificação terminológica dos códigos encontrados na etapa anterior, onde os recursos FHIR foram identificados. Para esta fase, o modelo se restringirá às terminologias da SNOMED, utilizando o Snowstorm

como servidor de terminologias. Nesse sentido, a camada terminológica é representada na Figura 14.

Figura 14 – Camada terminológica.



Fonte: Elaborado pelo autor.

Como mostra a Figura 14, o processo de identificação terminológica é baseado em requisições feitas via API ao servidor Snowstorm, com objetivo de transformar os recursos extraídos, na etapa anterior, em recursos enriquecidos com os códigos corretos da SNOMED. Cada recurso identificado é submetido à consulta ao servidor Snowstorm, onde os termos correspondentes na SNOMED são identificados e vinculados ao recurso. Esta etapa não apenas assegura que cada termo e conceito utilizado seja alinhado com a terminologia padrão da SNOMED, promovendo consistência e precisão na interoperabilidade, mas também representa um enriquecimento significativo dos dados.

O enriquecimento dos dados de saúde, proporcionado por esta etapa, amplia a utilidade dos recursos FHIR. Ao adicionar códigos SNOMED CT precisos, proporciona-se um nível de detalhe e de contextualização que vai além do escopo original dos dados, aumentando, dessa forma, a eficácia na comunicação entre diferentes sistemas de saúde. Este processo fortalece a interoperabilidade e enriquece a base de dados disponível para análises clínicas e pesquisa, oferecendo aos profissionais da saúde informações mais detalhadas e contextuais. Portanto, a

adoção dessa camada de enriquecimento é importante para facilitar uma compreensão mais profunda das condições de saúde dos pacientes.

A Tabela 8 exemplifica um recurso FHIR extraída na camada anterior (identificação FHIR) depois de passar pela camada de identificação terminológica:

Tabela 8 – Exemplo de identificação terminológica em recurso FHIR.

Recurso FHIR sem terminologia	**Recurso FHIR com terminologia identificada**

<pre>"resourceType": "AllergyIntolerance",
"patient": {
"reference": "Patient/example",
"display": "Paciente"
},
"substance": {
"coding": [
{
"system": "http://example.org/substance",
"code": "PastaDeAmendoim",
"display": "Pasta de Amendoim"
}
]
},
"reaction": [
{
"manifestation": [
{
"coding": [
{
"system": "http://example.org/symptom",
"code": "Desconforto",
"display": "Desconforto"
}
]
},
{
"coding": [
{
"system": "http://example.org/symptom",
"code": "Prurido",
"display": "Prurido"
}
]
},
{
"coding": [
{
"system": "http://example.org/symptom",
"code": "SinaisDeUrticaria",
"display": "Sinais de Urticária"
}
]
},
{
"coding": [
{
"system": "http://example.org/symptom",
"code": "InchacoLabial",
"display": "Inchaço Labial"
}
]</pre>
	<pre>"resourceType": "AllergyIntolerance",
	"patient": {
	"reference": "Patient/example",
	"display": "Patient"
	},
	"substance": {
	"coding": [
	{
	"system": "http://snomed.info/sct",
	"code": "**91935009**",
	"display": **"Peanut Allergy"**
	}
	]
	},
	"reaction": [
	{
	"manifestation": [
	{
	"coding": [
	{
	"system": "http://snomed.info/sct",
	"code": "**22253000**",
	"display": **"Pain"**
	}
	]
	},
	{
	"coding": [
	{
	"system": "http://snomed.info/sct",
	"code": "**26870002**",
	"display": **"Pruritus"**
	}
	]
	},
	{
	"coding": [
	{
	"system": "http://snomed.info/sct",
	"code": "**247255002**",
	"display": **"Urticaria"**
	}
	]
	},
	{
	"coding": [
	{
	"system": "http://snomed.info/sct",
	"code": "**44169009**",
	"display": **"Lip Swelling"**
	}
	]</pre>

Fonte: Elaborado pelo autor.

7.3.1 SNOMED CT

A SNOMED CT é uma terminologia clínica internacionalmente reconhecida, sendo um recurso preciso e abrangente para representação sistemática de termos clínicos. Ela permite um registro detalhado e codificado de dados clínicos, facilitando a troca de informações de saúde entre diferentes sistemas e plataformas (Cornet; De Keizer, 2008). Além disso, a SNOMED CT é fundamental no avanço da medicina baseada em evidências e no suporte à tomada de decisões clínicas, promovendo uma comunicação clara e precisa entre sistemas de informação em saúde e profissionais da área.

A adoção da SNOMED CT, entretanto, na prática clínica cotidiana, enfrenta desafios significativos, principalmente, relacionados à sua integração em larga escala. Para superar esses obstáculos, foi desenvolvido o Modelo Conceitual Legível por Máquina (MRCM), da SNOMED CT, aprovado em 2007 pelo IHTSDO.

A primeira versão de rascunho do MRCM, publicada em 2009, incluiu 234 restrições, como domínio, alcance e restrições de cardinalidade, marcando um passo significativo no desenvolvimento dessa ferramenta (SNOMED *International*, 2024). Nos anos seguintes, ocorreram avanços importantes, como a adoção do formato de distribuição padrão para a SNOMED CT, em 2012, e a publicação da primeira versão oficial da Linguagem de Restrição de Expressão, em 2015. Estes desenvolvimentos permitiram que o MRCM fosse aprimorado de forma mais padronizada e consistente.

O documento do MRCM define e descreve o design estrutural do modelo, abordando o contexto e as considerações levadas em conta durante seu desenvolvimento. Ele especifica regras, que definem domínio, alcance, cardinalidade e agrupabilidade de cada atributo, no modelo conceitual da SNOMED CT, representando um avanço significativo na gestão e aplicação de terminologias clínicas na área da saúde.

O público-alvo deste documento inclui centros nacionais de liberação da SNOMED, desenvolvedores de terminologia da SNOMED CT, e designers e desenvolvedores de sistemas de registros eletrônicos de saúde, entre outros, ressaltando a abrangência e a relevância do MRCM no campo da informática em saúde (SNOMED *International*, 2024).

A seguir, a Tabela 9 demonstra aplicações de uso da SNOMED em diversas áreas:

Tabela 9 – Uso da SNOMED CT em diversas áreas da saúde.

Área de Aplicação	Descrição	Benefícios
EMR	Uso da SNOMED CT para codificar informações clínicas, como diagnósticos, procedimentos, e condições de saúde.	Permite uma representação detalhada e codificada de dados clínicos, facilitando a troca de informações de saúde entre diferentes sistemas.
Pesquisa Clínica	Aplicação em estudos e pesquisas para categorizar e analisar dados clínicos de forma padronizada.	Facilita a análise de dados, comparação de resultados entre diferentes estudos e promove a medicina baseada em evidências.
Gestão em Saúde	Uso para análise de dados em saúde pública, gestão hospitalar e análise de desempenho de serviços de saúde.	Promove uma gestão de saúde mais eficiente, com análises precisas e decisões baseadas em dados confiáveis.
Farmacovigilância	Emprego na identificação e análise de eventos adversos relacionados a medicamentos.	Contribui para a segurança do paciente e monitoramento eficaz de medicamentos no mercado.
Interoperabilidade	Facilita a integração e comunicação entre diferentes sistemas de informação em saúde.	Promove a comunicação clara e precisa entre sistemas, otimizando a integração e gestão da informação clínica.
Apoio à Decisão Clínica	Utilizado para desenvolver sistemas de apoio à decisão clínica baseados em regras e lógicas codificadas.	Auxilia profissionais de saúde na tomada de decisões informadas, melhorando a qualidade do atendimento ao paciente.
Codificação e Classificação	Uso para codificar e classificar informações clínicas em sistemas de informação em saúde.	Assegura consistência e precisão na representação de informações clínicas, facilitando consultas e análises.

Fonte: Adaptado de SNOMED *INTERNATIONAL*.

A Tabela 9 destaca a versatilidade da SNOMED CT, em diferentes áreas da saúde, sublinhando seu papel crucial na padronização, gestão, e análise de informações clínicas.

Para uso efetivo e reconhecimento de códigos SNOMED, é comum o uso de servidores de terminologias em sistemas de informação. Neste sentido, o servidor de terminologia Snowstorm é um componente essencial para a manipulação e exploração de termos e conceitos do SNOMED CT. Esta ferramenta permite uma interação refinada e uma busca eficiente dentro do vasto conjunto de termos clínicos do SNOMED CT, tornando-se um recurso inestimável para profissionais de saúde e desenvolvedores na busca por informações padronizadas e unificadas em saúde

(Wassing, 2020). O Snowstorm é crucial para a superação de barreiras relacionadas à semântica e classificações não uniformes em registros médicos eletrônicos, favorecendo uma comunicação mais precisa e clara entre diferentes sistemas de informação em saúde.

7.3.2 Servidor de terminologias Snowstorm

O Snowstorm é um servidor de terminologia especializado e de código aberto, com suporte específico para SNOMED CT, desenvolvido e mantido pelo SNOMED *International*. Ele se destaca pelo desempenho e escalabilidade empresarial. Como servidor de terminologia focado no SNOMED CT, o Snowstorm implementa funcionalidades específicas para esta nomenclatura, incluindo o *Uniform Resource Identifier* (URI), do SNOMED CT, filtros, propriedades, conjuntos de valores e mapas de conceitos implícitos (IHTSDO, 2023).

O suporte para o *Snowstorm* é fornecido pela comunidade através do repositório do *GitHub*, já que o SNOMED *International* não oferece suporte comercial. O produto possui duas APIs principais: a FHIR API, que implementa o módulo de terminologia e suporta diversos sistemas de código, e a API Especialista em SNOMED CT, que suporta a gestão de sistemas de código SNOMED CT e a criação de edições SNOMED CT. Além disso, é a API do servidor de terminologia para a Plataforma de Autoria SNOMED *International*, usada para a manutenção da edição internacional do SNOMED CT e outras edições nacionais.

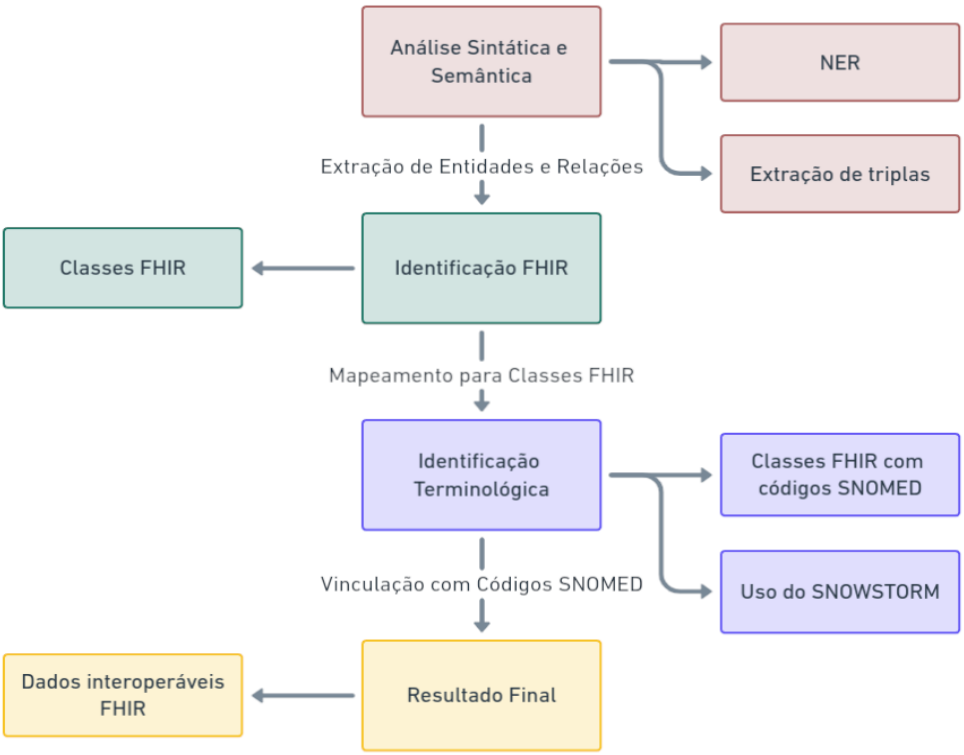
Para implementação, o uso de um serviço de terminologia como o *Snowstorm*, juntamente com uma interface padrão como FHIR API, permite uma integração flexível de aplicações e acesso aos recursos terminológicos avançados, como resolução de termos comuns em códigos oficiais e precisos da SNOMED. A documentação do *Snowstorm* inclui guias para configuração, uso das APIs, autoria SNOMED CT; e as contribuições ao projeto podem ser feitas por meio do repositório do projeto.

7.4 Validação do modelo

A representação final do FHIR-FLOW contempla as etapas de Extração de entidades, mapeamento de recursos FHIR, vinculação de códigos SNOMED CT, e

recursos FHIR finais devidamente classificados. É possível observar na Figura 15 o fluxo geral do modelo.

Figura 15 – Fluxo geral do modelo.



Fonte: Elaborado pelo autor.

A Figura 15 mostra uma visão geral para o mapeamento de dados de saúde para FHIR, estruturado em três etapas interligadas, cada uma com seus respectivos processos e artefatos. Na primeira etapa, Análise Sintática e Semântica, dados estruturados e não estruturados de EHR são analisados, utilizando técnicas como Extração de Sentenças e Reconhecimento de Entidades, gerando triplas que identificam relações entre entidades. Essas triplas são levadas à segunda etapa, Identificação FHIR, onde são mapeadas para recursos FHIR correspondentes através de modelos de linguagem avançados como o GPT-4. Na terceira e última etapa, Identificação Terminológica, aos recursos FHIR são alinhadas com códigos SNOMED CT corretos via consultas ao servidor Snowstorm, garantindo precisão terminológica.

Com o FHIR-FLOW implementado, a validação desempenha um papel essencial na avaliação da eficácia e confiabilidade da arquitetura de informação proposta. Ela foi conduzida com diversos objetivos em mente, com foco em assegurar

que o modelo atenda aos mais altos padrões de qualidade e cumpra sua finalidade de forma eficiente. A seguir estão os principais objetivos da validação:

- Avaliação da Conformidade FHIR: verificar se o modelo está em conformidade com o padrão FHIR, tanto em termos de estrutura, quanto de semântica dos dados. Isso envolve a validação da estrutura dos recursos FHIR que estão sendo utilizados e a conformidade com os perfis FHIR pertinentes.
- Integridade e Consistência dos Dados: é essencial garantir que os dados fornecidos pelo Hospital Sírio Libanês estejam completos, íntegros e coerentes. São verificações para identificar valores ausentes, valores extremos e quaisquer discrepâncias nos dados.

7.4.1 Avaliação da conformidade FHIR

No requisito de avaliação FHIR, foram feitos testes unitários com sentenças geradas sinteticamente pelo LLM, onde o intuito era avaliar a precisão de identificação dos recursos e a classificação terminológica correta, mediante a cenários distintos e aleatórios dentro dos cuidados de saúde, contemplando 3 amostras de:

- Anamneses;
- Receituário;
- Reação alérgica.

As tabelas 10, 11 e 12 foram construídas representando os recursos principais com destaque na cor azul, os atributos mapeados com a cor cinza e os valores atribuídos com a cor verde. Para os erros do modelo, foram usados destaques na cor vermelha.

Tabela 10 – Anamneses geradas para validação de conformidade FHIR.

Contexto Clínico	Amostra	Resultado	
Anamnese	Paciente de 45 anos, sexo masculino, relata cansaço frequente e dor no peito esporádica, principalmente após esforço físico. Refere histórico familiar de doenças cardíacas	“**resourceType**": "**Patient**", "**id**": "patient-example", "gender": "*male*", "birthDate": "*1975-01-01*" }, { "**resourceType**": "**Condition**", "clinicalStatus": { "coding": [{"system": " *http://snomed.info/sct* ", "code": "active"}] }, "code": { "coding": [{"system": "*http://snomed.info/sct*", "code": " *80449002* ", "display": *" Muscle fatigue (finding)* "}] }, "subject": {"reference": "Patient/patient-example"} },”	“**resourceType**": "**Condition**", "clinicalStatus": { "coding": [{"system": " *http://snomed.info/sct",* "code": "active"}]}, "code": { "coding": [{"system": "*http://snomed.info/sct*", "code": " *29857009* ", "display": *" Chest pain (finding)* "}] }, "subject": {"reference": "Patient/patient-example"} }, { "**resourceType**": "**FamilyMemberHistory**", "id": "example", "patient": { "reference": "Patient/patient-example" }, "condition": [{ "code": { "coding": [{ "system": "*http://snomed.info/sct*", "code": "*56265001* ", "display": "Heart disease (disorder)" }] } }] }
Anamnese	Mulher de 30 anos apresenta queixas de dores abdominais recorrentes, especialmente após as refeições. Relata episódios de diarreia e constipação alternados.	“**resourceType**": "**Patient**", "text": { "status": "generated" }, "gender": "*female*", "birthDate": "*1991-01-01*" },{ "**resourceType**": "**Condition**", "text": { "status": "generated" }, "code": { "coding": [{ "system": "*http://snomed.info/sct*", "code": "*21522001*", "display": "*Recurrent abdominal pain*" }, { "system": "*http://snomed.info/sct*", "code": "*56731001*", "display": "*Diarrhea*" },	{ "system": "*http://snomed.info/sct*", "code": "*236069006*", "display": "*Constipation*" }] }, "category": [{ "coding": [{ "system": " *http://snomed.info/sct*", "code": "*complaint*", "display": "*Complaint*" }] }], "clinicalStatus": { "coding": [{ "system": " *http://snomed.info/sct* ", "code": "*active*", "display": "*Active*" }] }, "subject": { "reference": "Patient/example" }}]

Fonte: Elaborado pelo Autor.

Conforme observado na tabela 10, no primeiro exemplo, um paciente masculino de 45 anos apresenta cansaço e dor no peito. O FHIR-FLOW corretamente identificou esses sintomas e o histórico familiar de doenças cardíacas, mapeando-os para os recursos FHIR apropriadas como *Condition* e *FamilyMemberHistory*. Isso reflete a capacidade do modelo de capturar informações detalhadas sobre o estado de saúde e histórico familiar do paciente. No segundo caso, uma mulher de 30 anos com queixas de dor abdominal, diarreia e constipação, após as refeições, foi processada

pelo modelo. Ele acertou ao mapear essas queixas para múltiplas condições *Condition* no FHIR, indicando a presença de dor abdominal recorrente, diarreia e constipação. Este exemplo mostrou a habilidade do modelo em lidar com múltiplos sintomas e traduzi-los em termos clínicos padronizados no FHIR.

Tabela 11 – Receituários gerados para validação de conformidade FHIR

Contexto Clínico	Amostra	Resultado	

Receituário	Paciente: Homem, 45 anos
	Medicação: Atenolol 50mg
	Posologia: Tomar 1 comprimido por dia pela manhã
	Indicação: Controle da hipertensão e prevenção de dor no peito.
		"**resources**": [
		{
		"**resourceType**": "**Patient**",
		"gender": "*male*",
		"birthDate": "*1976-01-01*"
		}, {
		"**resourceType**": "**MedicationStatement**",
		"medicationCodeableConcept": {
		"coding": [{
		"system": "http://hl7.org/fhir/sid/ndc",
		"code": "Atenolol 50mg"
		}]
		},
		"dosage": [{
		"text": "*Tomar 1 comprimido por dia pela manhã*"
		}]
		},
			{
			"**resourceType**": "**Condition**",
			"code": {
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": "*38341003*",
			"display": "*Hypertension*"
			}] },
			"clinicalStatus": {
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": "active"
			}] } }, {
			"**resourceType**": "**Condition**",
			"code": {
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": "*22298006*",
			"display": "*Chest pain*"
			}]}

Receituário	Para a paciente de 30 anos que apresenta dores abdominais e sintomas de intestino irritável, prescrevo Mebeverina 200mg.
	Ela deve tomar uma cápsula três vezes ao dia, preferencialmente 20 minutos antes de cada refeição principal.
		{
		"**resourceType**": "**Patient**",
		"id": "example",
		"active": true,
		"birthDate": "*1991-04-20*",
		"gender": "*female*"
		},
		{
		"**resourceType**": "**Condition**",
		"code": {
		"coding": [{
		"system": "*http://snomed.info/sct*",
		"code": " *21522001*",
		"display": " *Abdominal pain (finding)* "
		}]
		},
		"subject": {
		"reference": "Patient/example"
		},
		"clinicalStatus": {
		"coding": [{
		"system": " *http://snomed.info/sct*
		",
		"code" : "active"
		}]
		}
		},
			"**resourceType**": "**Condition**",
			"code": {
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": "10743008",
			"display": "*Irritable bowel syndrome (disorder)*"
			}]}, "subject": {
			"reference": "Patient/example"
			},
			"clinicalStatus": {
			"coding": [{
			"system":
			"http://terminology.hl7.org/CodeSystem/condition-clinical",
			"code": "active"
			}] } }, {
			"**resourceType**": "**MedicationStatement**",
			"dosage": [{"method": {
			"coding": [{"system":
			"*http://snomed.info/sct*",
			"code" : "*26663004*",
			"display" : "*Oral*"
			}] }, "doseAndRate" : [{
			"doseQuantity" : {
			"value": *200*,
			"unit": "*mg*",
			"medicationCodeableConcept": {
			"coding": [{
			"system": **"http://www.whocc.no/atc",**
			"code": " *419830007* ",
			"display": " *Mebeverine (substance)* "
			}]}, "subject": {
			"reference": "Patient/example"
			} }

Fonte: Elaborado pelo Autor.

Conforme observado na tabela 11, no primeiro caso, apesar do erro na utilização da terminologia "http://hl7.org/fhir/sid/ndc" em vez de SNOMED CT, o FHIR-FLOW acertou ao mapear outros aspectos do caso para recursos FHIR corretos, como *Patient*, para detalhes do paciente, e *Condition*, para os sintomas relatados. Isto demonstra necessidade de ajuste no modelo, mas também a eficiência dele em interpretar e classificar informações clínicas em estruturas FHIR relevantes.

No segundo caso, o modelo cometeu erros na escolha das terminologias, mas conseguiu acertar nas inferências relacionadas à condição clínica do paciente, mapeando corretamente para os recursos FHIR, como *Patient*, para informações do paciente, e *Condition*, para os sintomas de dor abdominal, diarreia e constipação. Isso indica que, apesar dos desafios na seleção de terminologia, o modelo mantém sua capacidade de discernir e categorizar efetivamente dados clínicos em estruturas FHIR apropriadas.

Tabela 12 – Dados de alergia gerados para validação de conformidade FHIR.

Contexto Clínico	**Amostra**	**Resultado**	

História de alergia	Um paciente de 28 anos veio à consulta relatando uma reação alérgica incomum.
	Após consumir frutos do mar, especialmente camarão, ele notou uma erupção cutânea avermelhada e pruriginosa em seu corpo.
		"**resourceType**": "**Patient**",
		"identifier": [{
		"use": "usual",
		"type": {
		"coding": [{
		"system": " *http://snomed.info/sct* ",
		"code" : "MR"
		}]
		},
			"**resourceType**":
			"**AllergyIntolerance**",
			"code": {
			"coding": [{
			"system": "*http://snomed.info/sct",*
			"code": "227493005",
			"display": "Cashew nuts"
			}] },
			"criticality": "*high*",
			"reaction": [{
			"manifestation": [{
			"concept": {
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": "*39579001*",
			"display": " *Anaphylaxis (disorder)* "
			}] } }],
			"note": [{
			"text": "*The patient reports that the onset of urticaria was within 15 minutes after eating cashew*s."
			}],
			"severity" : "*severe*"
			}]
			}

História de alergia	Paciente: 35 anos
	
	Data de Identificação da Alergia: 10/03/2024
	
	Substância Alergênica: Amendoim
	
	Tipo de Reação Alérgica: Urticária Edema de Glote Dificuldade Respiratória
	
	Gravidade: Grave
	
	Último Episódio: 15/03/2024
	
	Informações Adicionais:
	O paciente teve sua primeira reação alérgica a amendoins aos 30 anos, caracterizada por urticária e dificuldade respiratória
		"**resourceType**": "**Condition**",
		"id": "condition-example",
		"clinicalStatus":
		"coding": [{
		"system": " *http://snomed.info/sct* ",
		"code": "active"
		}]}, "verificationStatus": {
		"coding": [
		{"system": " *http://snomed.info/sct* ", "code": "*confirmed*"
		}] },
		"code": {
		"coding": [
		{"system": "*http://snomed.info/sct",*
		"code": "*227493005*",
		"display": "*Cashew nuts*"
		}],
		"text": "*Alergia a amendoins*"
		},
		"severity": {
		"coding": [
		{
		"system": "*http://snomed.info/sct*",
		"code": "*24484000*",
		"display": "*Severe*"
		}
		]
		}
			"**resourceType**":
			"**AllergyIntolerance**",
			"id": "allergyintolerance-example",
			"category": ["*food*"],
			"criticality": "*high*",
			"code": {
			"coding": [{
			"system": "*http://snomed.info/sct",*
			"code": "*227493005*",
			"display": "*Cashew nut (substance)"*
			}] },
			"patient": {
			"reference": "Patient/example"
			},
			"reaction": [{
			"manifestation": [{
			"coding": [{
			"system": "*http://snomed.info/sct*",
			"code": " *126485001* ",
			"display": " *Urticaria (disorder)* "
			}, {
			"system": "*http://snomed.info/sct*",
			"code": " *111278002* ",
			"display": *" Edema of glottis (disorder)*
			"}, {
			"severity": "*Severe*"
			}] },

Fonte: Elaborado pelo Autor.

Conforme observado na tabela 12, no caso de um paciente de 28 anos com reação alérgica após consumo de camarão, o modelo identificou corretamente a situação clínica, mas cometeu um erro ao mapear para uma alergia a *Cashew nuts* (castanhas de caju) em vez de camarão. Embora o recurso FHIR *AllergyIntolerance*

tenha sido acertadamente selecionada, com detalhes sobre a reação e a criticidade, o erro na identificação específica do alérgeno indica uma limitação no reconhecimento e mapeamento correto de alérgenos específicos.

No segundo caso, um paciente de 35 anos, com histórico de reação alérgica a amendoins, o modelo foi mais bem-sucedido. Ele identificou corretamente a alergia a amendoins, classificando-a na categoria *food* dentro do recurso *AllergyIntolerance*. As reações específicas, como urticária e edema de glote, foram mapeadas corretamente, assim como a gravidade da condição. Este resultado demonstra a capacidade do modelo de processar e categorizar com precisão informações complexas sobre alergias em estruturas FHIR.

Ao avaliar a performance geral do modelo de interoperabilidade de dados de saúde para FHIR, observou-se um desempenho notável em diversas frentes. Na Tabela 10, relacionada às anamneses, o modelo exibiu uma alta precisão ao interpretar e mapear adequadamente informações clínicas complexas para os recursos FHIR corretos, como *Patient* e *Condition*. Contudo, nas tabelas subsequentes, que envolviam dados de alergia, o modelo apresentou inconsistências. Embora tenha acertado em muitas inferências, como na identificação correta de reações alérgicas e suas classificações em *AllergyIntolerance*, ocorreram erros na terminologia utilizada e na identificação exata de alérgenos específicos. Esses resultados indicam uma capacidade robusta do modelo em processar e mapear informações clínicas, e, também, destacam a necessidade de refinamentos, especialmente na precisão da terminologia e no reconhecimento de alérgenos, para assegurar sua aplicabilidade e eficácia em ambientes clínicos reais.

Para extrair um percentual de acurácia, foi adotada uma metodologia detalhada para calcular as métricas de acerto e erro. Cada atributo mapeado para um recurso FHIR, em cada entrada, foi considerado como uma inferência separada. Assumindo que cada entrada nas tabelas continha múltiplas inferências, estimou-se que cada entrada continha aproximadamente 10 inferências, incluindo a correta identificação de recursos e atributos como *Patient*, *Condition*, e *AllergyIntolerance*, e a precisão na codificação SNOMED CT.

Com base nessa estimativa, o FHIR-FLOW alcançou um desempenho de cerca de 66,67% de acertos nas inferências ao longo das três tabelas analisadas. No entanto, também se observou um índice de erro de aproximadamente 33,33%,

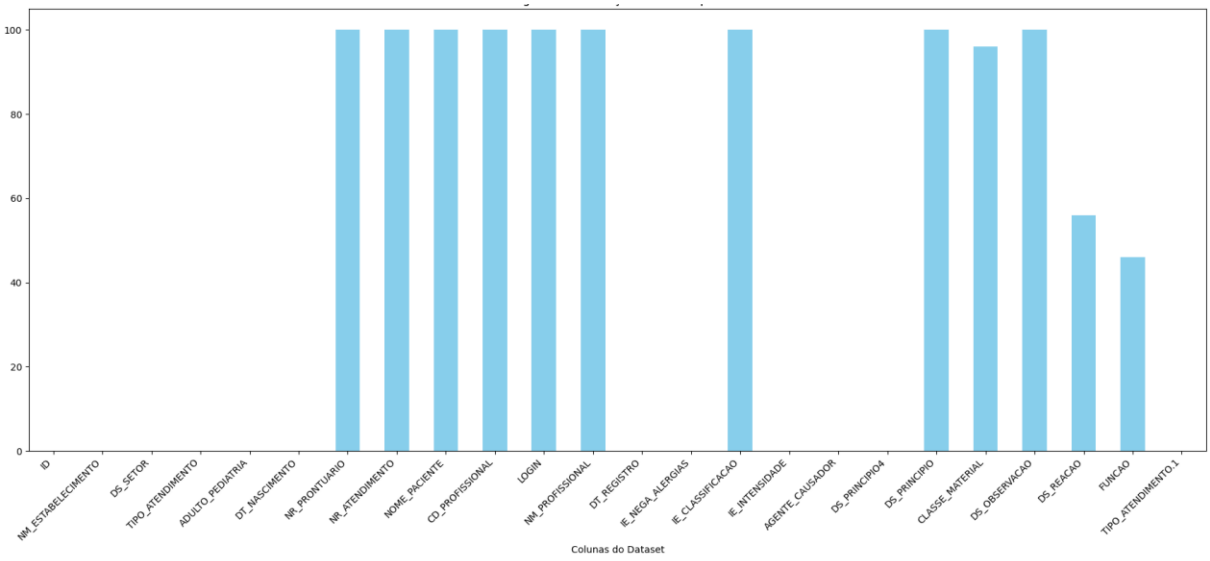
refletindo áreas que requerem melhorias, especialmente, na precisão das terminologias e na identificação exata de detalhes clínicos.

7.4.2 Integridade e consistência dos dados

Para a validação do modelo, mediante aos dados do HSL, foi realizada uma amostragem com 50 registros aleatórios do *dataset* anonimizado fornecido pelo hospital. Dos 50 registros, dos quais:

- 14 tinham dados de “AGENTE CAUSADOR” descrito como “OUTROS”
- 28 tinham dados de ausentes de “REAÇÃO ALÉRGICA”
- 23 tinham dados de “FUNÇÃO” do profissional atendente ausentes

Gráfico 1 – Análise de presença e ausência de dados.



Fonte: Elaborado pelo Autor.

A análise da presença e ausência de dados, visualizada por meio do gráfico de porcentagem de dados ausentes por coluna, constitui uma etapa importante na preparação e na avaliação da qualidade do *dataset* utilizado neste estudo, pois esta visualização não apenas destaca as colunas que requerem atenção especial, devido à alta incidência de dados ausentes, mas também serve como revelação das deficiências na coleta e no registro de informações de saúde, oferecendo uma transparência de aspectos que podem afetar a eficácia do modelo.

O modelo, que conseguiu identificar e gerar o recurso FHIR, correspondente à 23 registros, no qual classifica os recursos como *AllergyIntolerance*., Ao total, obteve-

se um percentual de classificação de 40%, mas considerando que 90% dos registros tinham todas as informações completas de reação alérgica, conforme pode ser visualizado na amostra de dados, disponibilizada na Tabela 13.

Tabela 13 – Amostra de dados.

ID de registro	Descrição da Alergia	Informações de Reação	Data da Reação	Histórico da Exposição	Observações Adicionais
001	Alergia a penicilina	Rash, prurido	2022-05-15	Primeira prescrição	Paciente relatou inchaço
002	Intolerância à lactose	Distensão abdominal, diarreia	2022-07-20	Consumo de laticínios	Sintomas persistiram por 48 horas
003	Alergia a amendoim	Anafilaxia	2022-08-11	Ingestão acidental	Admitido no pronto-socorro
004	Sensibilidade ao glúten	Fadiga, dores articulares	2022-09-05	Dieta contendo glúten	Melhora após ajuste na dieta
005	Alergia ao látex	Urticária, dificuldade respiratória	2022-10-02	Contato com luvas de látex	Sintomas aliviados com antihistamínico
006	Reação adversa a ibuprofeno	Náusea, vômito	2022-11-15	Uso para dor de cabeça	Evitado em futuras prescrições
007	Alergia a picada de abelha	Inchaço, vermelhidão, dor no local	2022-12-22	Picada durante caminhada	Aplicação de epinefrina necessária
008	Dermatite de contato por níquel	Erupção cutânea, coceira	2023-01-08	Uso de bijuterias	Troca por acessórios hipoalergênicos
009	Alergia a aspirina	Urticária, angioedema	2023-02-14	Administração para febre	Monitoramento de reações em futuro uso

Fonte: Elaborado pelo Autor.

A Tabela 13 reflete uma variedade de informações coletadas dos registros de saúde, abrangendo não apenas a descrição da alergia e as informações de reação, mas também detalhes importantes como data da reação, histórico de exposição e quaisquer observações adicionais feitas pelo profissional de saúde. Esses dados ilustram a complexidade e riqueza das informações contidas nos prontuários eletrônicos, destacando o desafio inicial de extrair e interpretar esses dados de maneira eficiente e precisa antes da aplicação de qualquer modelo de processamento. A Tabela 14 disponibiliza a amostra de dados validados.

Tabela 14 – Amostra de dados validados.

ID do Registro	Tipo de Alergia	Reação	Gravidade	Confirmação	Recurso FHIR
001	Medicamento	Rash, prurido	Moderada	Sim	*AllergyIntolerance*
003	Alimento	Anafilaxia	Severa	Sim	*AllergyIntolerance*
005	Material	Urticária	Severa	Sim	*AllergyIntolerance*
007	Inseto	Inchaço, vermelhidão	Moderada	Sim	*AllergyIntolerance*
009	Medicamento	Urticária, angioedema	Moderada	Sim	*AllergyIntolerance*
011	Alimento	Diarreia, vômito	Leve	Sim	*AllergyIntolerance*
013	Ambiente	Congestão nasal	Leve	Sim	*AllergyIntolerance*
015	Químico	Dermatite	Moderada	Sim	*AllergyIntolerance*
017	Animal	Espirros, coceira	Leve	Sim	*AllergyIntolerance*

Fonte: Elaborado pelo autor.

A Tabela 14 acima exibe uma seleção de dez registros, que foram eficientemente processados e classificados pelo modelo FHIR-FLOW, ilustrando a capacidade do sistema em interpretar a complexidade dos dados de saúde e mapeá-los, com precisão, para o recurso FHIR *AllergyIntolerance*. Cada linha na tabela demonstra a conversão bem-sucedida de descrições detalhadas de alergias em dados estruturados e padronizados, incluindo tipo de alergia, natureza da reação, gravidade, confirmação da condição e classificação no formato FHIR. Em contrapartida, a Tabela 15 apresenta os dados não validados.

Tabela 15 – Dados não validados.

ID do Registro	Descrição da Alergia	Informações de Reação	Dados Faltantes
002	Intolerância à lactose	Distensão abdominal, diarreia	Gravidade, Confirmação
004	Sensibilidade ao glúten	Fadiga, dores articulares	Tipo de Alergia, Gravidade
006	Reação adversa a ibuprofeno	Náusea, vômito	Gravidade, Confirmação
008	Dermatite de contato por níquel	Erupção cutânea, coceira	Confirmação
010	Sensibilidade a perfumes e fragrâncias	Dor de cabeça, irritação nos olhos	Tipo de Alergia, Gravidade
012	Alergia a corantes alimentares	Urticária	Gravidade, Confirmação
014	Reação adversa a conservantes	Distúrbios gastrointestinais	Tipo de Alergia, Gravidade, Confirmação
016	Alergia a poeira doméstica	Espirros, congestão nasal	Confirmação
018	Sensibilidade a mudanças climáticas	Mal-estar, fadiga	Tipo de Alergia, Gravidade

Fonte: Elaborado pelo Autor.

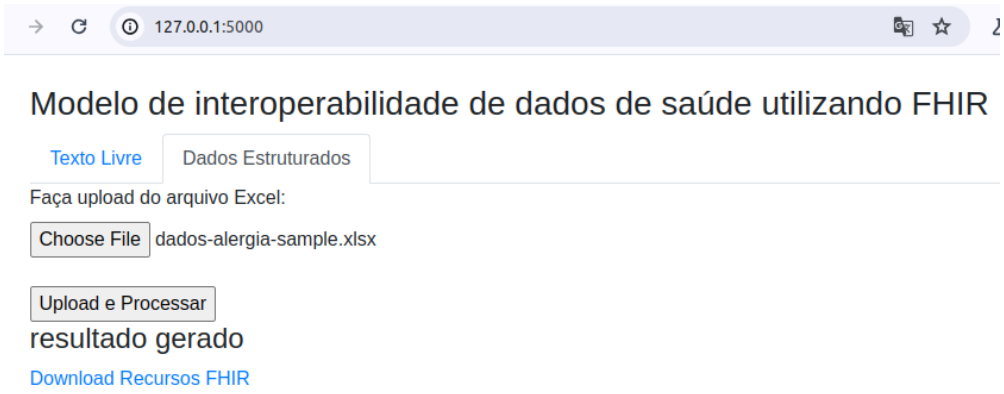
A Tabela 15 foca nos registros que não foram validados pelo modelo, evidenciando lacunas nas informações, que impediram a classificação precisa em um recurso FHIR. Estes registros apresentam desafios como dados incompletos, ausência de confirmação de diagnóstico ou falta de clareza na descrição das alergias e reações. A tabela ainda demonstra o cenário enfrentado pelo FHIR-FLOW, quando as entradas não cumprem os requisitos mínimos para uma análise conclusiva, ressaltando, dessa forma, a importância de dados completos e precisos para a eficácia do modelo. A Figura 16 demonstra a validação de texto livre como amostragem.

Figura 16 – Amostra de validação de texto livre com a aplicação atuando.

Fonte: Elaborado pelo autor.

A Figura 16 exemplifica o processo de validação dos dados com a tela da aplicação do modelo implementado. Na aba “Texto livre”, é possível digitar amostras de cenários clínicos de forma livre, para o modelo processar conforme as camadas descritas ao longo da seção. Na segunda aba, é possível fazer *upload* de um arquivo em formato *Excel*, representando o esquema dos dados estruturados.

Figura 17 – Tela de validação de dados estruturados e *download* do resultado.



Fonte: Elaborado pelo autor.

A Figura 17 mostra a tela para inserção de arquivo e validação de dados estruturados. Ela possui como opção realizar o *download* com resultados dos recursos FHIR, em formato JSON.

8 CONSIDERAÇÕES FINAIS

Esta pesquisa, com o objetivo principal de desenvolver um modelo de interoperabilidade de dados de saúde, usando o padrão FHIR, seguiu uma metodologia estruturada, que incluiu análise de literatura, desenvolvimento de modelo arquitetônico e implementação de um protótipo. A introdução do estudo detalhou a construção de um *corpus* documental, iluminando tendências e desafios na interoperabilidade de dados de saúde. O estudo, em suas seções iniciais, aprofundou-se na revisão da literatura, identificando contribuições e lacunas relevantes.

Ao explorar o padrão FHIR, a pesquisa destacou a necessidade de um modelo arquitetônico eficiente, para gerir a complexidade e diversidade dos dados de saúde. A construção do modelo proposto e o desenvolvimento de um *software* foram detalhados nas seções. Os resultados obtidos demonstraram que a aplicação de um modelo baseado no padrão FHIR pode significativamente melhorar a interoperabilidade dos dados, facilitando um fluxo de informações mais eficiente e aumentando a eficácia dos processos de saúde.

Do ponto de vista crítico, os resultados contribuem significativamente para a Ciência da Informação e para a área da saúde. Eles demonstraram o impacto positivo de um modelo arquitetural para interoperabilidade aprimorado com uso de tecnologias, como *Machine Learning* e LLMs, associados aos padrões internacionalmente reconhecidos, como o FHIR. Contudo, existem limitações e desafios, como a escalabilidade do modelo em diferentes contextos de saúde e sua precisão terminológica, que evidenciam a necessidade de pesquisas futuras para superar estas limitações e maximizar o potencial do padrão FHIR.

8.1 Conclusões

A presente pesquisa iluminou um caminho rumo à superação dos obstáculos persistentes na interoperabilidade de sistemas de saúde, com o padrão FHIR, emergindo não apenas como uma ferramenta, mas como um habilitador para este avanço. A implementação do modelo FHIR-FLOW, central para este estudo, demonstrou não só a viabilidade, mas a necessidade de uma arquitetura sofisticada para lidar com a complexa questão de dados de saúde. Este modelo facilitou um fluxo de informações mais integrado e eficiente entre sistemas de saúde dispares, e

também estabeleceu um novo padrão para a qualidade e a acessibilidade do atendimento ao paciente.

Em uma era marcada pela crescente demanda por serviços de saúde personalizados e baseados em evidências, o FHIR-FLOW surge como uma resposta promissora aos desafios de longa data associados à interoperabilidade dos dados. Ao integrar tecnologias avançadas como *Machine Learning* e Modelos de Linguagem Grande, este estudo transcende a tradicional barreira entre os silos de dados, promovendo uma compreensão mais profunda e um uso mais eficaz das informações de saúde.

Além disso, esta pesquisa destacou a sinergia entre a Ciência da Informação e a tecnologia da informação, dentro do ecossistema da saúde. A interdisciplinaridade não foi apenas uma abordagem metodológica, mas uma necessidade para abraçar completamente a complexidade inerente à interoperabilidade dos dados de saúde. A colaboração entre estas disciplinas enriqueceu a pesquisa e forneceu *insights* valiosos sobre como os dados de saúde podem ser estruturados, interpretados e aplicados para melhorar os resultados dos pacientes.

Em suma, o FHIR-FLOW representa um passo significativo na direção de uma saúde digital mais conectada e interoperável. Este estudo não apenas contribuiu para o corpo acadêmico da Ciência da Informação e tecnologia da saúde, mas também ofereceu um modelo replicável, que pode ser adaptado e implementado em diversas configurações de saúde, para promover uma colaboração mais efetiva e um atendimento ao paciente mais informado e personalizado.

8.2 Contribuições

A pesquisa contribui significativamente para compreensão da interoperabilidade de dados de saúde e do papel do padrão FHIR. As contribuições deste estudo transcendem a simples apresentação de um modelo teórico, adentrando na aplicabilidade prática e nos benefícios práticos para o ecossistema de saúde como um todo. Dentre as contribuições chave, destacam-se:

- Desenvolvimento e Validação de um Modelo Inovador: o modelo FHIR-FLOW representa um marco na jornada rumo a uma interoperabilidade aprimorada, oferecendo um guia concreto para a integração de sistemas de saúde multifacetados. Este modelo demonstra a aplicação prática do padrão FHIR e incorpora tecnologias emergentes como *Machine Learning* e Processamento

de Linguagem Natural para superar desafios de interpretação e integração de dados.

- Análise Crítica e Construtiva: a exploração detalhada da situação atual da interoperabilidade nos sistemas de saúde, e dos desafios prevalentes, forneceram um pano de fundo crítico para a compreensão da complexidade inerente à integração de dados de saúde. Este estudo não se limitou a identificar problemas, ele avançou propondo soluções inovadoras e viáveis, enraizadas em rigoroso embasamento teórico e prático.
- Proposição de Diretrizes Estratégicas: através da proposição de diretrizes e recomendações estratégicas para a implementação eficaz do padrão FHIR, este estudo atua como um farol, orientando instituições de saúde, desenvolvedores de *software* e formuladores de políticas na adoção de práticas, que maximizem a eficiência e eficácia dos dados de saúde.
- Implicações Práticas e Estratégicas: o protótipo desenvolvido ilustra a viabilidade do modelo FHIR-FLOW, e serve como um testamento das potencialidades práticas da interoperabilidade avançada em sistemas de saúde. Esse modelo prático oferece *insights* valiosos, não apenas para implementações futuras, mas também, como um estudo de caso, sobre o impacto direto da interoperabilidade na qualidade do atendimento ao paciente.

Ao iluminar o caminho para uma integração mais eficaz de dados de saúde, este estudo marca um ponto de inflexão na forma como os dados de saúde são compartilhados, gerenciados e utilizados. A implementação do padrão FHIR, conforme explorado neste trabalho, é um passo crítico na transformação digital da saúde, promovendo uma era onde a informação é fluida, acessível e, significativamente, mais impactante para o atendimento ao paciente.

O legado do modelo FHIR-FLOW, portanto, reside não apenas em sua contribuição para a interoperabilidade dos dados de saúde, mas também como um habilitador para futuras pesquisas e desenvolvimentos na interseção da Ciência da Informação, tecnologia da saúde e cuidado ao paciente.

REFERÊNCIAS

AYAZ, M.; PASHA, M. F.; ALZAHRANI, M. Y.; BUDIARTO, R.; STIAWAN, D. The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. **JMIR Medical Informatics**, v. 9, n. 7, 2021.

BALCH, J. A.; RUPPERT, M. M.; LOFTUS, T. J.; GUAN, Z.; REN, Y.; UPCHURCH, G. R.; OZRAZGAT-BASLANTI, T.; RASHIDI, P.; BIHORAC, A. Machine Learning–Enabled Clinical Information Systems Using Fast Healthcare Interoperability Resources Data Standards: Scoping Review. **JMIR Medical Informatics**, v. 11, 2023.

BOUH, M.; HOSSAIN, F.; AHMED, A. A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format. *In:* INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES FOR AGEING WELL AND E-HEALTH, 9., 2023, Praga, **Proceedings** […]**,** Praga: Science and Technology Publications Lda, 2023.

BRANDT, M. B. **Modelagem da informação legislativa: arquitetura da informação para o processo legislativo brasileiro**. 2020. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2020.

BRAUNSTEIN, M. Healthcare in the age of interoperability: the promise of fast healthcare interoperability resources. **IEEE pulse**, v. 9, n. 6, p. 24-27, 2018.

BRASIL. **Portaria nº 2.073**. 2011. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073_31_08_2011.html. Acesso em: 20 abr. 2024.

CASELI, H.; FREITAS, C.; VIOLA, R. Processamento de Linguagem Natural. *In:* BRAZILIAN SYMPOSIUM ON DATA BASES, 37., 2022, **Short courses of the 37th Brazilian Symposium on Data Bases** […], Búzios, 2022.

CHATTERJEE, A.; PAHARI, N.; PRINZ, A. HL7 FHIR with SNOMED-CT to achieve semantic and structural interoperability in personal health data: a proof-of-concept study. **Sensors**, v. 22, n. 10, p. 3756, 2022.

CORNET, R.; DE KEIZER, N. Forty years of SNOMED: a literature review. **BMC Medical Informatics and Decision Making**, v. 8, Suppl 1:S2, out. 2008.

DE LA CRUZ, S.; GONZÁLEZ, I.; GARCÍA, T.; ROSARIO, M. Alergias alimentarias**:** Importancia del control de alérgenos en alimentos. **Nutrición Clínica Dietética Hospitalaria**, v. 38, n. 1, p. 142-148, 2018.

FOREMAN, C. *et al.* Penicilin allergy labeling on medical alert jewelry. **Annals of Allergy, Asthma and Immunology**, v. 124, n. 6, p. 627-629, 2020.

FRANSSON, S.; MOSBECH, H.; KAPPEL, M.; HJORTLUND, J.; POULPSEN, L.; KVISSELGAARD, A. D.; GARVEY, L. H. The importance of prolonged provocation in

drug allergy: results from a Danish allergy clinic. **The Journal of Allergy and Clinical Immunology:** In Practice, v. 5, n. 5, p. 1394-1401, 2017.

GHASSEMI, M.; NAUMANN, T.; SCHULAM, P.; BEAM, A. L.; CHEN, I. Y.; RANGANATH, R. A review of challenges and opportunities in machine learning for health. **AMIA Summits on Translational Science Proceedings**, v. 2020, p. 191, 2020.

GRABER, M.L.; BYRNE C.; JOHNSTON, D. The impact of electronic health records on diagnosis. **Diagnosis** (Berl), v. 4, n. 4, p. 211-223, 2017. DOI 10.1515/dx-2017-0012.

HAMM, R.; KNOOP, S. E.; SCHWARZ, P.; BLOCK, A. D.; DAVIS, W. L. Harmonizing clinical terminologies: driving interoperability in healthcare. **Studies in health technology and informatics**, v. 129, n. 1, p. 660-663, 2007.

HL7 FHIR. **AllergyIntolerance resource.** 2021. Disponível em: https://www.hl7.org/fhir/allergyintolerance.html. Acesso em: 20 abr, 2024.

HUNG, C. Y.; CHEN, H.Y.; WEE, L. J.; LIN, C. H.; LEE, C. C. Deriving a novel health index using a large-scale population base electronic health record with deep networks. *In:* 42nd ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE & BIOLOGY SOCIETY (EMBC), 42., 2020, **Proceedings** […], 2020.

HUGGING FACE. **Documentação do modelo BERT**. Disponível em: https://huggingface.co/docs/transformers/model_doc/bert. Acesso em: 4 de janeiro de 2024.

IQBAL, A. M.; SHEPHERD, M.; ABIDI, S. S. R. An ontology-based electronic medical record for chronic disease management. *In:* 44TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 44., 2011, Hawaii. **44th Hawaii International Conference on System Sciences**. Hawai: IEEE, 2011. p. 1-10.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685-695, 2021.

JAVAID, M.; HALEEM, A.; SINGH, R. P.; SUMAN, R.; RAB, S. Significance of machine learning in healthcare: Features, pillars and applications. **International Journal of Intelligent Networks**, v. 3, p. 58-73, 2022.

KADDOUR, J.; HARRIS, J.; MOZES, M.; BRADLEY, H.; RAILEANU, R.; MCHARDY, R. Challenges and applications of large language models. **arXiv:2307.10169**, jul. 2023.

LIU, X.; CHEN, H.; XIA, W. Overview of named entity recognition. **Journal of Contemporary Educational Research**, v. 6, n. 5, p. 65-68, 2022.

NARDON, F. B.; MOURA JUNIOR, L. de A. **Compartilhamento de conhecimento em saúde utilizando ontologias e bancos de dados dedutivos**. **2003.**Universidade de São Paulo, São Paulo, 2003.

NOUMEIR**,** R. Active Learning of the HL7 Medical Standard. **Journal of Digital Imagin**, v. 32, p. 354–361, 2019. Disponível em: https://doi.org/10.1007/s10278-018-0134-3. Acesso em: 20 abr. 2024.

OLIVEIRA, B. S. N.; RÊGO, L. G. C. do; PERES, L.; SILVA, T. L. C. da; MACÊDO, J. A. F. de. Processamento de Linguagem Natural via Aprendizagem Profunda. *In*: ALMEIDA, E. S.; SANTORO, F. M. (orgs.). **41ª Jornada de Atualização em Informática**. Porto Alegre: SBC, 2022.

OPENAI. **Visão geral dos modelos**. Disponível em: https://platform.openai.com/docs/models/overview. Acesso em: dez. 2023.

Organização Internacional de Terminologia em Saúde (IHTSDO). Disponível em: https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=81268943. Acesso em: 20 abr. 2024.

SNOWSTORM. Disponível em: https://github.com/IHTSDO/snowstorm. Acesso em: dez. 2023.

PADOVANI, D. **Um método adaptativo para análise sintática do Português Brasileiro**. 2022. Tese (Doutorado em Sistemas Digitais) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2022.

PATEL, A. A.; ARASANIPALAI, A. U. **Applied Natural Language Processing in the Enterprise:** teaching machines to read, write, and understand. Califórnia: O'Reilly Media, 2021.

PETRY, K.; TOMAZELLA, L. H. B.; ANDRADE, R.; VON WANGENHEIM, A. Utilização do Padrão HL7 para Interoperabilidade em Sistemas Legados na Área de Saúde. *In*: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 11., 2008. **Anais** [...]. 2008.

PICKLER, M. E. V. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, v. 12, p. 65-83, 2007.

PIMENTA, N.; CHAVES, A.; SOUSA, R.; ABELHA, A.; PEIXOTO, H. Interoperability of Clinical Data through FHIR: A review. **Procedia Computer Science**, v. 220, p. 856-861, 2023.

PINE, K. H. The qualculative dimensiono f healthcare data interoperability. **Health Informatics Journal**, v. 25, n. 3, p. 536-548, 2019.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Editora Feevale, 2013.

ROEHRS, A. COSTA, C. A. da; RIGHI, R. da R.; RIGO, S. J.; WICHMAN, M. H. Toward a model for personal health record interoperability. **IEEE Journal of Biomedical and Health Informatics**, v. 23, n. 2, p. 867-873, 2018.

SARIPALLE, R.; RUNYAN, C.; RUSSELL, M. Using HL7 FHIR to achieve interoperability in patient health record. **Journal of biomedical informatics**, v. 94, p. 103-188, 2019.

SHERMAN, R. Align Metadata and Business Initiatives. **DM Review**, New York, v. 16, n. 1, p. 50, jan. 2006.

SINGHAL, K. *et al***.** Large language models encode clinical knowledge. **Nature**, v. 620, n. 7972, p. 172-180, 2023.

SNOMED INTERNATIONAL. **SNOMED Introduction**. Disponível em: https://confluence.ihtsdotools.org/display/DOCMRCM/1.+Introduction. Acesso em: 3 jan. 2024.

STAN, O.; MICLEA, L. Local EHR management based on FHIR. *In*: IEEE INTERNATIONAL CONFERENCE ON AUTOMATION, QUALITY AND TESTING, ROBOTICS (AQTR), 2018. **IEEE Proceedings** […], 2018. p. 1-5.

THIRU, K.; HASSEY, A.; SULLIVAN, F. Systematic review of scope and quality of electronic patient record data in primary care. **BMJ**, v. 326, n. 7398, p. 1070, 2003.

TIERNEY, M. J.; PAGELER, N. M.; KAHANA, M.; PANTALEONI, J. L.; LONGHURST, C. A. Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions. **Acad Med.,** v. 88, n. 6, p. 748-752, 2013. DOI: 10.1097/ACM.0b013e3182905ceb. Acesso em: 20 abr. 2024.

WASSING, D. **Evaluation of terminology servers for use with Snomed CT**. 2020. Dissertação (Mestrado em Ciência da Computação) – Linkõping University, Suécia, 2020

WEGNER, P. Interoperability. **ACM Computing Surveys** (CSUR), v. 28, n. 1, p. 285-287, 1996.

YANG, T.; HE, Y.; YANG, N. Named Entity Recognition of Medical Text Based on the Deep Neural Network. **Journal of Healthcare Engineering**, v. 2022, p. 1-10, 2022.