



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  
2023

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:** Dr. Leonardo Castro Botega

Marília  
2024

## RESUMO

A evolução dos registros eletrônicos de saúde trouxe benefícios, mas enfrenta um desafio: a interoperabilidade dos dados, onde a falta de padrões unificados impede a troca eficaz de informações entre sistemas de saúde. Esta pesquisa foca na criação de um modelo de arquitetura de informação para a interoperabilidade de dados de saúde, utilizando o padrão FHIR (*Fast Healthcare Interoperability Resources*). O método envolveu análise de dados de alergia de prontuários eletrônicos do Hospital Sírio Libanês e revisão bibliográfica sobre interoperabilidade para embasamento e construção do modelo FHIR-FLOW, que utiliza em suas camadas técnicas de aprendizado de máquina e auxílio de *Modelo de Linguagem Grande* (LLM). Os resultados destacam que a implementação do padrão FHIR aliada a métodos de análise de dados não só melhora a interoperabilidade, mas também facilita a comunicação entre sistemas de saúde distintos. Este avanço é essencial para aprimorar a gestão de saúde, indicando um caminho efetivo para resolver as dificuldades na comunicação de dados de saúde, um passo fundamental para o avanço dos cuidados de saúde na era digital.

**PALAVRAS-CHAVE:** Arquitetura da informação; FHIR;

**ABSTRACT**

The evolution of electronic health records has brought benefits, but faces a challenge: data interoperability, where the lack of unified standards prevents the effective exchange of information between healthcare systems. This research focuses on creating an information architecture model for healthcare data interoperability, using the FHIR (Fast Healthcare Interoperability Resources) standard. The method involved analysis of allergy data from electronic medical records at Hospital Sírio Libanês and bibliographic review on interoperability to support and build the FHIR-FLOW model, which uses machine learning techniques and the aid of Large Language Models in its layers. The results highlight that the implementation of the FHIR standard combined with data analysis methods not only improves interoperability, but also facilitates communication between different health systems. This advancement is essential to improving healthcare management, indicating an effective path to solving difficulties in communicating healthcare data, a fundamental step towards advancing healthcare in the digital era.

**KEYWORDS:** Information Architecture;FHIR

### LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema passo a passo da metodologia.....	16
Figura 2 - Mapa conceitual do ambiente informacional de um prontuário eletrônico e sua semântica .....	21
Figura 3 - Visão de acesso a recursos na arquitetura FHIR .....	29
Figura 4 - Diagrama UML parcial do recurso FHIR Patient .....	30
Figura 5 - Mapa conceitual da arquitetura do padrão FHIR .....	32
Figura 6 - Etapas do modelo BERT.....	41
Figura 7 - Machine Learning na área da saúde.....	47
Figura 8 - Representação, recuperação e acesso a informação de dados clínicos .....	51
Figura 9 - Mapa conceitual dos atributos do dataset de alergia do HSL .....	54
Figura 10 - Camada de identificação FHIR .....	70
Figura 11 - Camada Terminológica .....	72
Figura 12 - Visão geral do modelo .....	77
Figura 13 - Amostra de validação de texto livre com a aplicação rodando .....	85
Figura 14 - Tela de validação de dados estruturados e download do resultado .....	86

**LISTA DE ABREVIATURAS E SIGLAS**

<b>API</b>	Application Programming Interface
<b>AI</b>	Arquitetura da Informação
<b>CI</b>	Ciência da Informação
<b>EBM</b>	Evidence Based Medicine
<b>EMR</b>	Eletronic Medical Records
<b>ERP</b>	Enterprise Resource Planning
<b>FHIR</b>	Fast Healthcare Interoperability Resources
<b>HL7</b>	Health Level 7
<b>HSL</b>	Hospital Sírio Libanês
<b>HTTP</b>	Hypertext transfer Protocol
<b>IA</b>	Inteligência Artificial
<b>ICD</b>	International Classification of Diseases
<b>JSON</b>	Javascript Object Notation
<b>LLM</b>	Modelo de Linguagem Grande
<b>LOINC</b>	Logical Observation Identifiers Names and Codes
<b>ML</b>	Machine Learning
<b>MIMIC</b>	Medical Information Mart for Intensive Care
<b>MRCM</b>	Modelo Conceitual Legível por Máquina
<b>NER</b>	Named Entity Recognition
<b>NLP</b>	Natural Processing Language
<b>PEP</b>	Prontuário Eletrônico do Paciente
<b>PLN</b>	Processamento de Linguagem Natural
<b>REST</b>	Representational Transfer State
<b>SNOMED CT</b>	Systematized Nomenclature of Medicine – Clinical Terms
<b>TISS</b>	Troca de Informação para Saúde Suplementar
<b>UML</b>	Unified Modeling Language
<b>URI</b>	Uniform Resource Identifier

<b>VA</b>	Veteran's Administration
<b>XML</b>	Extensible Markup Language

## LISTA DE TABELAS

Tabela 1 - Comparativo entre trabalhos correlatos.....	27
Tabela 2 - Principais tipos de Machine Learning e correlação com a CI .....	36
Tabela 3 - Exemplo de prompt de chamada para o GPT3.5-turbo .....	45
Tabela 4 - Amostra original dos dados de alergia do HSL.....	53
Tabela 5 - Metadados de negócio gerados com interoperabilidade com o padrão FHIR.....	60
Tabela 6 - Exemplo de extração de triplas a partir de fontes de dados .....	67
Tabela 7 - Exemplo de identificação de classes FHIR através de triplas.....	71
Tabela 8 - Exemplo de identificação terminológica em classe FHIR .....	73
Tabela 9 - Uso da SNOMED-CT em diversas áreas da saúde.....	75
Tabela 10 - Anamneses geradas para validação de conformidade FHIR.....	79
Tabela 11 - Receituários gerados para validação de conformidade FHIR.....	81
Tabela 12 - Dados de alergia gerados para validação de conformidade FHIR .....	83



## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>11</b>
1.1 Problema de Pesquisa .....	12
1.2 Justificativa .....	13
1.3 Metodologia.....	14
1.4 Estrutura da Pesquisa .....	17
<b>2 PANORAMA DOS DADOS DE SAÚDE.....</b>	<b>19</b>
2.1 Interoperabilidade de Dados de Saúde .....	23
<b>3 INTEROPERABILIDADE DE DADOS DE SAÚDE: TRABALHOS CORRELATOS .....</b>	<b>25</b>
<b>4 O PADRÃO FHIR .....</b>	<b>28</b>
4.1 Arquitetura Geral do FHIR.....	28
4.2 Recursos FHIR.....	29
<b>5 MACHINE LEARNING NA ÁREA DA SAÚDE.....</b>	<b>34</b>
5.1 Potencial e aplicações .....	35
5.2 Processamento de Linguagem Natural .....	38
5.4 Modelos de Aprendizado de Máquina em PLN: BERT .....	40
5.5 Modelos de Linguagem Grande (LLMs).....	42
5.6 Desafios na Integração de ML e Saúde .....	46
<b>6 DADOS DE ALERGIA DO HOSPITAL SÍRIO LIBANÊS: ANÁLISE COM METADADOS DE NEGÓCIO .....</b>	<b>49</b>
6.2 Importância de dados de alergia .....	49
6.2 Aplicação de metodologia de arquitetura da informação.....	50
<b>7 FHIRFLOW: MODELO DE MAPEAMENTO DE DADOS DE SAÚDE PARA FHIR .....</b>	<b>62</b>
7.1 Camada Sintática e Semântica.....	65
7.2 Camada Identificação FHIR .....	69
7.3 Camada Terminológica .....	72
7.3.1 A SNOMED-CT .....	74
7.3.2 O servidor de terminologias Snowstorm .....	76
7.4 Validação do modelo.....	77
7.4.1 Avaliação da conformidade FHIR.....	78
7.4.2 Integridade e consistência dos dados .....	85
<b>8 CONSIDERAÇÕES FINAIS.....</b>	<b>87</b>
8.1 Conclusões .....	87
8.2 Contribuições .....	88

**REFERÊNCIAS..... 89**

## 1 INTRODUÇÃO

Com o aumento de 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 dos dados através dos EMR (*Eletronic Medical Records*), e também, de acordo com Tierney (2013), na influência direta maneira como os profissionais ministram os cuidados aos e auxiliando no pensamento clínico crítico.

Mediante o número de instituições e softwares existentes, um dos desafios dos EMR é a padronização de arquitetura da informação, 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 também a existência de um grande volume de padrões e arquiteturas existentes, onde cada instituição tem dificuldade de escolha particular no momento de escolher a arquitetura que lhe trará um melhor custo-benefício (PETRY et al, 2008).

Neste sentido, a interoperabilidade dos dados de saúde, que por definição é 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 permita a troca de dados entre os diferentes sistemas, para permitir o aumento da capacidade de organização e recuperação dos dados, e, portanto, gerando 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, existem 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 entre outros. No Brasil, o TISS (Troca de Informação para Saúde Suplementar) 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 tanto clínicas quanto administrativas e financeiras, e permite o acompanhamento financeiro das operadoras de convênios médicos. Já o padrão FHIR (*Fast Healthcare Interoperability*

*Resources*) é desenvolvido pela HL7®( Health Level 7 International) e é um protocolo internacional para envio e recebimento de dados na área da saúde que contempla informações clínicas e administrativas, e vem de encontro com a necessidade de integração de dados na área da saúde para otimizar a pesquisa e 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, assegurando a fidedignidade da informação, simplificando e unificando a pesquisa e recuperação das informações. Inclusive no Brasil, há uma preocupação com a interoperabilidade de sistemas médicos e isso ficou evidente 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**

Considerando a importância do cuidado de saúde e a 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 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 também 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 não apenas dificulta o compartilhamento de informações médicas, mas também 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 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 têm 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, e esta lacuna na interoperabilidade impacta diretamente a agilidade e a qualidade dos serviços prestados, tornando necessário o desenvolvimento de soluções eficazes para essa questão.

O presente estudo justifica-se em várias frentes. Primeiramente, 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 é capaz de elucidar perspectivas relevantes para a elaboração de novas estratégias e políticas de gestão da informação em saúde, objetivando a promoção da interoperabilidade. Isso 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, então, abordar

a questão da interoperabilidade, com a 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, do ponto de vista de sua natureza, 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 seus objetivos trata-se de uma pesquisa exploratória e documental pois possui como finalidade proporcionar mais informações sobre o assunto a ser investigado, possibilitando a sua definição e o seu delineamento, e também faz 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, que foram fornecidos pelo Hospital Sírio Libanês.

A construção da arquitetura do modelo proposto foi iniciada com uma análise detalhada dos dados de alergia fornecidos pelo Hospital Sírio Libanês (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 assim 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 SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms), 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 a realização de testes e a validação dos protocolos do padrão. Este modelo, nomeado FHIR-FLOW, foi então 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 BRAPCI, IEEEExplore, entre outras, usando o termo "interoperabilidade saúde" e "health interoperability", respectivamente.

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 IEEEExplore, dos 17 artigos retornados, 4 foram pertinentes à proposta da pesquisa. O objeto de análise considerado neste estudo são os artigos científicos.

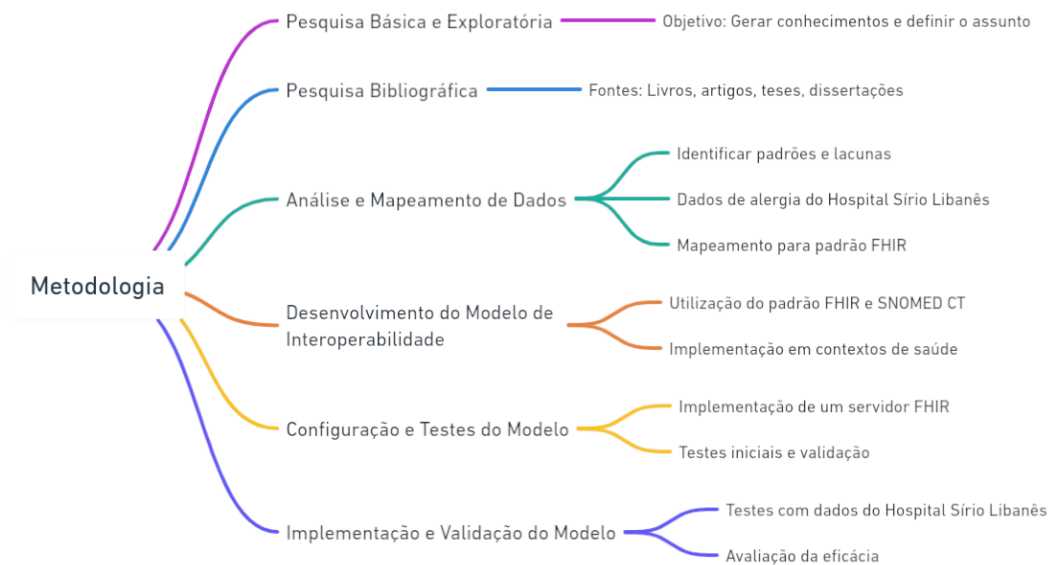
Com a literatura em mãos, os artigos selecionados foram submetidos a uma leitura detalhada, buscando destacar o problema de pesquisa, os objetivos, a metodologia, os resultados e as conclusões de cada trabalho. Este processo permitiu uma compreensão mais profunda dos desafios e avanços recentes no campo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

No aspecto prático 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, serão 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. Adicionalmente, 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 negócio que contribuiu estabelecendo uma conexão entre a teoria e a realidade do setor de saúde, e também permitiu, com os dados, a validação efetiva do modelo de interoperabilidade proposto.

Este estudo não se deteve na análise e revisão bibliográfica; ele avançou na implementação e validação do modelo de interoperabilidade desenvolvido. Esta fase 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 se concentrou apenas na interoperabilidade de dados de saúde existentes e nos desafios a ela 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.

*Figura 1 - Esquema 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. Em seguida, 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, além de expor os principais desafios e oportunidades relacionados ao tema em questão. É detalhado o problema de pesquisa, a justificativa e os objetivos gerais e específicos do estudo, e além disso, 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 do paciente, incluindo dados de origem exclusiva dos pacientes e dados provenientes de contatos assistenciais. Também é destacada 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. São apresentados 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 e de radiologia. 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, denominada “Interoperabilidade de dados de Saúde: Trabalhos Correlatos”, são apresentados os trabalhos correlatos que abordam a integração de dados de saúde. São mencionados estudos relevantes, como o de Karine et al. (2008), que propõe um modelo de interoperabilidade utilizando o padrão HL7, e o 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. Também é abordado o trabalho de Braunstein (2018), que discute os níveis de interoperabilidade desejados no contexto da saúde. 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 “O Padrão FHIR”, é abordado em maior detalhe o padrão FHIR, promovido pela HL7, e seu objetivo de determinar uma transferência representacional do estado (REST) para representar as entidades e procedimentos de saúde como recursos. É discutido o papel dos recursos no FHIR e como eles definem a estrutura e o conteúdo de informações transmitidas entre sistemas. É mencionada também a utilização de terminologias no FHIR, que vinculam os dados a vocabulários comuns, como SNOMED, LOINC (Logical Observation Identifiers Names and Codes) e ICD (International Classification of Diseases). 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* (ML), 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. É explorado o uso de algoritmos de ML na categorização de informações em registros médicos e na conversão de dados não estruturados em dados estruturados. São apresentadas aplicações do ML na medicina, como o diagnóstico auxiliado por computador e a personalização de tratamentos médicos e também ao final, mencionados os desafios e oportunidades do uso de ML na saúde.

Na seção 6, denominada “Dados de Alergia do Hospital Sírio Libanês: Análise com Metadados de Negócio”, é abordada uma aplicação prática de integração de dados e interoperabilidade no contexto do Hospital Sírio Libanês. 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 gerou *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 principal é abordar e solucionar os desafios de interoperabilidade de dados agindo como uma ponte, transformando dados brutos e heterogêneos em informações padronizadas e prontas para interoperabilidade. Também é explorada 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).

A extensão dos dados de saúde é 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 também resultados de exames, que subsidiam os profissionais nas tomadas de decisões de diagnósticos quanto nas evoluções posteriores, mediante a medicações e procedimentos. Atualmente, existem mais de 4000 testes de laboratório selecionáveis, e um número comparativamente desconcertante de opções de imagem como afirma Graber (2017).

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 nossa 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 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. Além disso, 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 contextualização dos dados, pois possibilitam a 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 um prontuário eletrônico e sua semântica

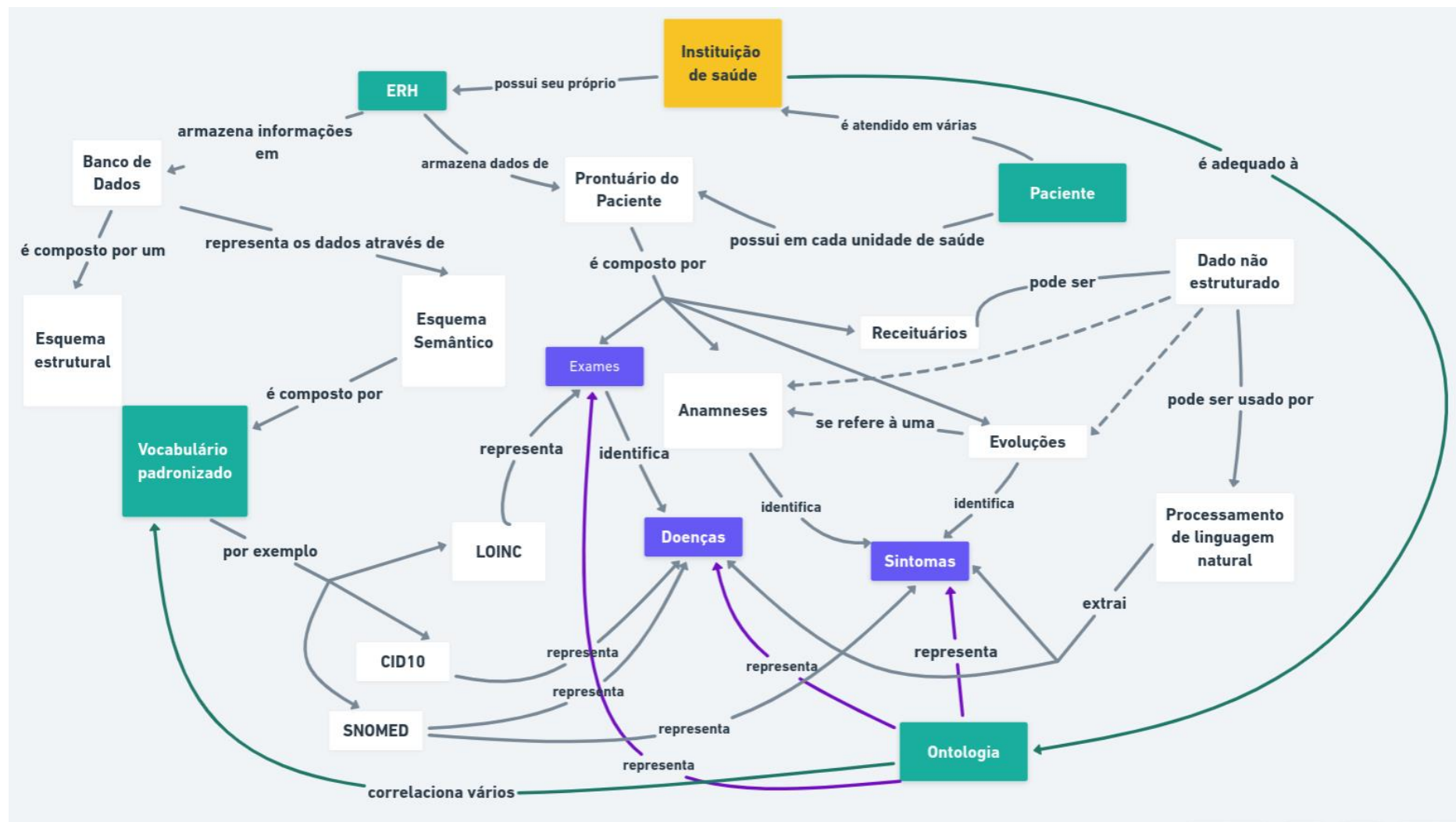


Figura adaptada de Iqbal et al., 2011

Na figura 2, pode-se observar um sistema de registros eletrônicos de saúde (ERH), como diversos componentes interagem de forma integrada. No coração deste sistema está o próprio ERH, que atua como um núcleo central armazenando e gerenciando informações essenciais de saúde. O ERH 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 e é estruturado em dois aspectos principais: o esquema estrutural, que define a arquitetura dos dados, e o esquema semântico, que interpreta o significado desses dados. As instituições de saúde, cada uma com seu próprio sistema de ERH, 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, que por conta disso é construído 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 garantem que os dados sejam compreensíveis e interoperáveis entre diferentes sistemas e são usados para representar exames, doenças e sintomas de maneira padronizada, facilitando a compreensão e o compartilhamento de informações. Além disso, parte dos dados nos prontuários, como receituários, anamneses e evoluções, são frequentemente não estruturados, e neste cenário entra em atuação o processamento de linguagem natural (PLN), 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, ajudando na organização e na interpretação dos dados de saúde, facilitando a identificação de padrões e conexões, como a relação entre exames específicos e as doenças que eles ajudam a diagnosticar. Em resumo, neste sistema de ERH, 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, e em um cenário ideal, permitindo 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. Pine (2019) afirma que “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:

- Troca de informações na gestão de consultórios, clínicas e hospitais e é especialmente útil para instituições que atuam em todos os níveis de atenção ao paciente, permitindo o rastreamento clínico do indivíduo nos serviços utilizados;
- Compartilhamento de dados do prontuário eletrônico do paciente (PEP) com segurança para aumento de evidências que subsidiam decisões clínicas;
- Disponibilização de resultados de exames laboratoriais e de radiologia, permitindo 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, contribuindo 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 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. Já 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 várias propostas de soluções e caminhos a serem adotados para realizar a interoperabilidade que se complementam, e padrões para trocas de 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 e aumenta a eficiência nos processos e até redução de custos. Na próxima serão apresentados trabalhos correlatos realizados no cenário de interoperabilidade de dados de saúde, apresentando 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 a 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 Karine et al (2008) introduziu um modelo baseado no padrão HL7, projetado para um servidor de troca de mensagens envolvendo 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, um aspecto que permaneceu obscuro foi a abordagem de mapeamento dos dados originais, que não foi detalhada 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 openEHR, HL7 FHIR e MIMIC-III. Um dos triunfos desse trabalho foi a eficaz implementação de técnicas de inteligência artificial e processamento de linguagem natural (NLP) 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.

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, evidenciando sua adoção por instituições de grande envergadura, como *Medicare* e *Veteran's Administration* (VA). No entanto, embora sua discussão tenha sido rica em insights, Braunstein 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. Propuseram a utilização do

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, *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.

Balch et al (2023) exploraram a aplicação e potencial dos sistemas de informação clínica habilitados para machine learning (ML) no contexto da transformação da entrega e pesquisa em saúde. Sublinharam a crescente integração do padrão de dados Fast Healthcare Interoperability Resources (FHIR) nesses sistemas, apesar das variações em métodos de implementação. A pesquisa revelou avanços notáveis, como o uso inovador de sistemas em nuvem, redes Bayesianas e estratégias de visualização, bem como técnicas de conversão de dados não estruturados para FHIR. Entretanto, identificaram-se limitações significativas: muitos sistemas avançados ainda enfrentam barreiras de interoperabilidade com prontuários eletrônicos de saúde e apresentam uma 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. Entretanto, embora o trabalho tenha abordado a promissora implementação do padrão FHIR para interoperabilidade de dados clínicos, baseou-se principalmente em uma revisão da literatura existente, não tendo as descobertas e conclusões validadas experimentalmente em ambientes práticos, o que limita a aplicabilidade direta das recomendações em cenários clínicos reais.

A seguir, uma tabela comparativa entre os trabalhos estudados:

Tabela 1 - Comparativo entre trabalhos correlatos

Autor(es) e Ano	Padrões Abordados	Pontos Fortes	Limitações	Uso de IA?
Karine 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)	openEHR, 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 O 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 o objetivo de fornecer um conjunto de recursos e APIs (*Application Programming Interface*) baseadas em REST (*Representational State Transfer*) para facilitar o acesso e utilização 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, tendo como ideia central construir um conjunto básico de recursos que, isoladamente ou combinados, atende a muitos usos comuns casos (STAN e 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 é sua capacidade de fornecer uma camada semântica, permitindo a associação dos dados de saúde a terminologias internacionais, como SNOMED, LOINC e ICD. Isso facilita a padronização e a 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 (*Representational State Transfer*) e 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 XML (*Extensible Markup Language*), JSON (*JavaScript Object Notation*) ou RDF (*Resource Description Framework*), e são acessíveis através de APIs (*Application Programming Interfaces*) baseadas em HTTP (*Hypertext Transfer Protocol*). 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 implementação do FHIR é 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 a 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 a recursos na arquitetura FHIR*

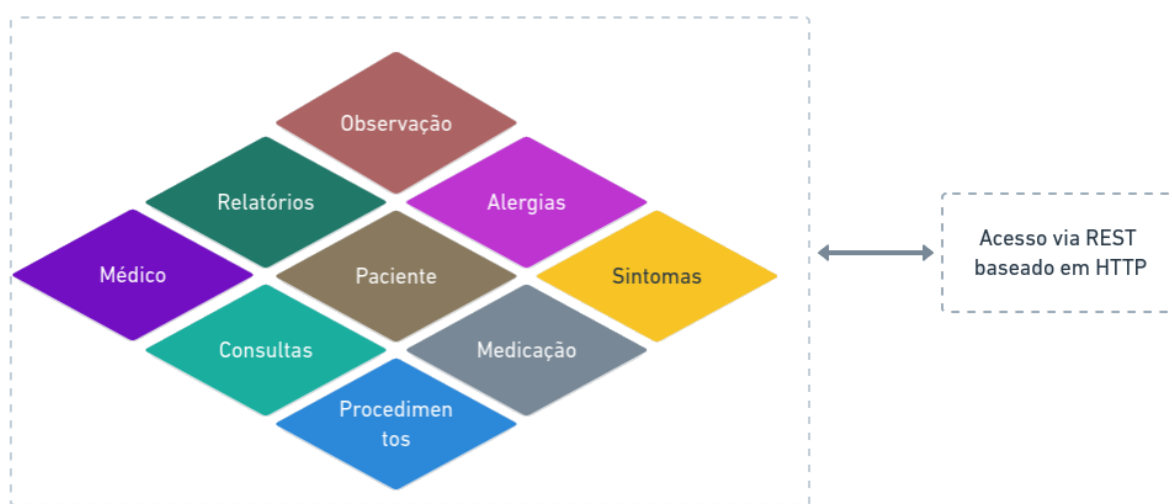


Figura adaptada de HL7 FHIR. (2021)

A Figura 3 representa o acesso aos recursos FHIR, que são os 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

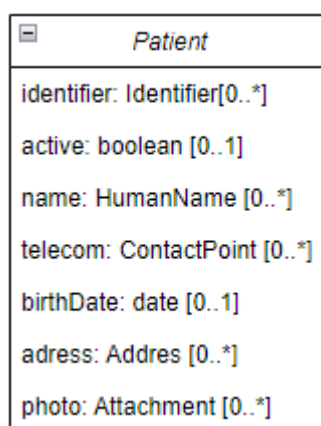


Figura adaptada de HL7 FHIR. (2021)

A Figura 4 apresenta um diagrama UML (*Unified Modeling Language*) representando a classe *Patient*, baseada no recurso de mesmo nome do FHIR. A classe inclui respectivamente os 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 a 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 os dados clínicos a terminologias internacionais já estabelecidas, como SNOMED, LOINC, ICD-9 e ICD-10. Essa integração permite o reaproveitamento de vocabulários existentes e amplamente utilizados, e também 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*

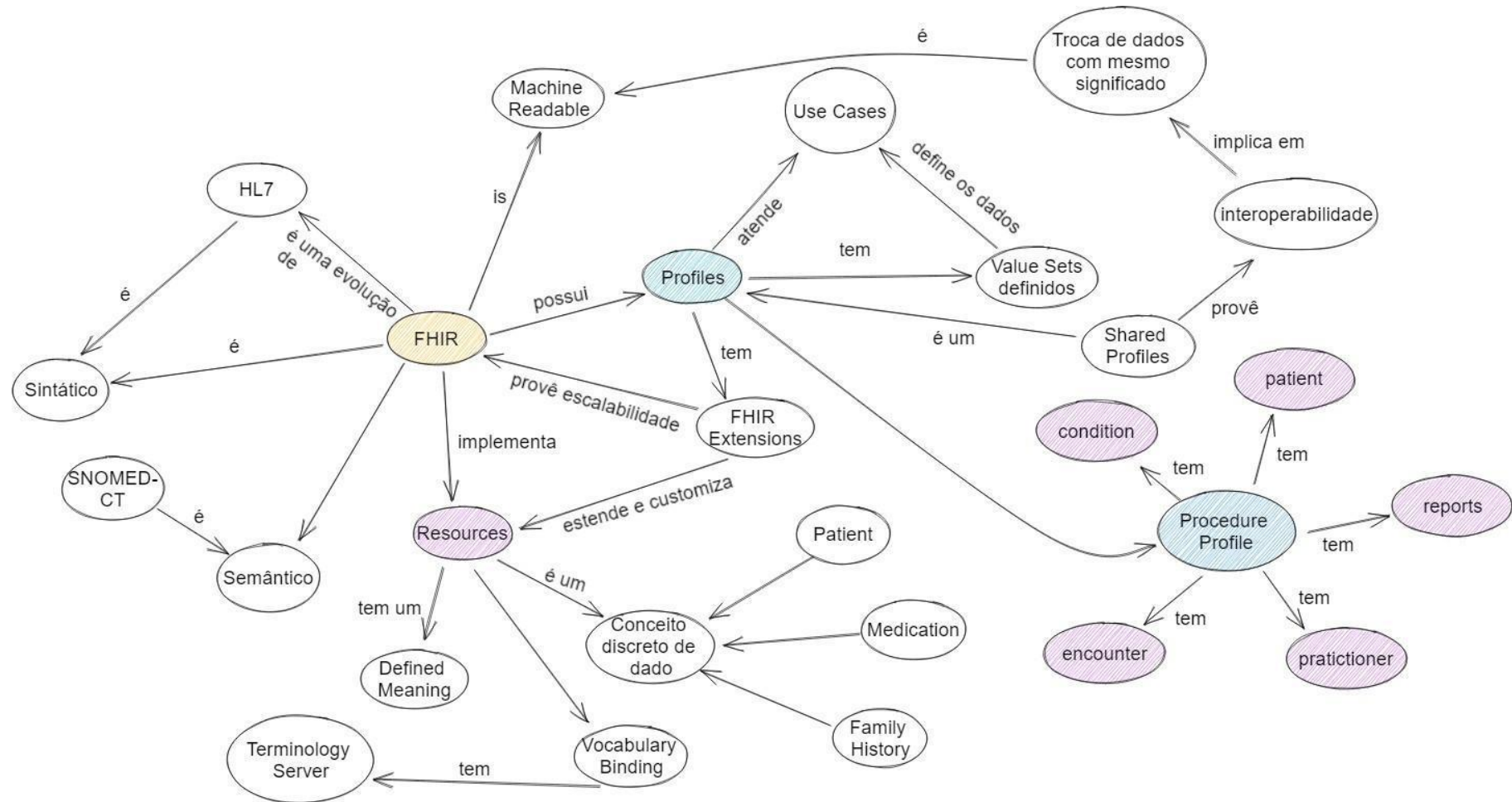


Figura adaptada de HL7 FHIR. (2021)



Na Figura 5 pode ser observada a arquitetura geral do FHIR , que representa uma evolução do padrão HL7 e é 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 as 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. Em resumo, 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 foi destacada a grande 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, o ML não só potencializa a análise clínica como também 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 à comunicação eficiente entre diferentes sistemas, dispositivos e plataformas.

Neste cenário, o *Machine Learning* surge como uma ferramenta transformadora. O ML é 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). Ele permite que sistemas melhorem automaticamente sua performance com a 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 a conversão de registros médicos em papel para formatos digitais. Algoritmos de ML podem ser treinados para reconhecer padrões complexos em dados de saúde.

Além disso o *Machine Learning* 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 não só tem gerado inovações significativas, mas também 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 é intimamente ligado à Ciência da Informação (CI) e desempenha um papel crucial na recuperação de informações. Além disso, 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, e a tabela 2 apresenta os principais tipos de *Machine Learning* e faz uma correção com conceitos fundamentais da Ciência da Informação (CI):

Tabela 2 - Principais tipos de Machine Learning e correlação com a CI

<b>Tipo de Aprendizado</b>	<b>Explicação</b>	<b>Correlação com a Ciência da Informação</b>
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 ciência da informação. 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.

Adaptado 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. Modelos de ML têm sido utilizados para detectar tumores em imagens de ressonância magnética e para identificar doenças oculares em imagens de retina, conforme Javaid et al. (2022) destacam. 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. Isso 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. Esses pilares incluem:

- **Previsão de Surto:** Utilização de algoritmos para antecipar e preparar-se para surtos de doenças, maximizando a prontidão e a resposta.
- **Descoberta e Fabricação de Medicamentos:** Acelerando a identificação de compostos terapêuticos e otimizando os processos de fabricação.
- **Modificação Comportamental:** Usando dados e análises para incentivar comportamentos saudáveis e mudanças de estilo de vida.
- **Registros de Saúde Inteligentes:** Aprimorando a gestão e análise dos prontuários eletrônicos para extração de insights clínicos.
- **Coleta de Dados Colaborativa:** Potencializando a coleta de informações de múltiplas fontes, tornando a pesquisa e o diagnóstico mais abrangentes.
- **Melhor Radioterapia:** Otimizando tratamentos de radioterapia para maior eficácia e minimização de efeitos colaterais.
- **Diagnóstico de Imagens Médicas:** Melhorando a precisão do diagnóstico por meio da análise automatizada de imagens, como raios-X e ressonância magnética.
- **Clínica e Pesquisa:** Facilitando 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. Isso não apenas aprimora a eficiência dos dados de saúde, mas também contribui para a interoperabilidade de dados, viabilizando a troca e utilização de informações em diferentes sistemas.

## 5.2 Processamento de Linguagem Natural

Dentro da Ciência da Informação, o Processamento de Linguagem Natural (PLN) é uma área interdisciplinar que integra conhecimentos da Ciência da Computação, Inteligência Artificial e Linguística. Segundo Caseli (2022), o PLN é essencial para a Ciência da Informação, fornecendo métodos e ferramentas para compreender, interpretar e gerar linguagem humana, e 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.

O PLN, após a etapa inicial de compreensão dos padrões linguísticos, conforme descrito por Caseli (2022), avança para a análise léxica. Essa fase consiste em decompor o texto em unidades menores, como palavras ou termos, denominados *tokens*. A análise léxica é essencial para identificar e isolar os elementos básicos da linguagem, e uma vez realizada a tokenização, segue-se a 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 a análise semântica. Esta fase tem como objetivo entender o significado das palavras e sentenças. Isso é alcançado mapeando palavras e frases a entidades em um domínio específico, 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 a identificação de sinônimos, antônimos e outras relações semânticas. Esta técnica é essencial para modelos de linguagem avançados como o BERT (Bidirecional Encoder Representations from Transformers), 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.

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, especialmente quando combinada com padrões de interoperabilidade como FHIR, é útil para construir sistemas de saúde integrados e coesos, pois garante 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 não só impulsiona a pesquisa clínica, como também revela inovações e direcionamentos para futuras pesquisas, contribuindo para a extração de conhecimento de dados brutos em informações que podem ser utilizadas.

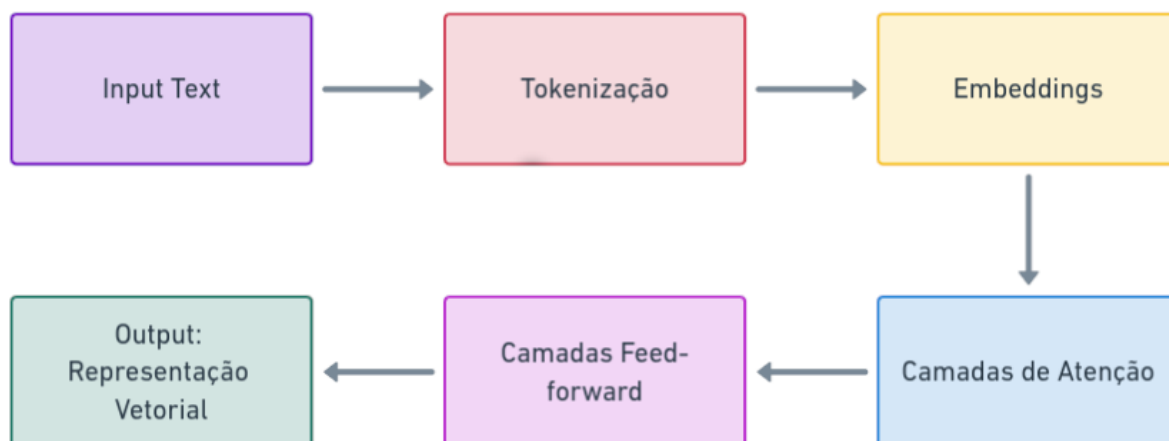
Considerando a relevância do PLN na transformação digital dos registros médicos e sua integração com padrões de interoperabilidade como FHIR, a próxima seção da tese se volta para a exploração detalhada do BERT (Bidirectional Encoder Representations from Transformers), como uma das tecnologias de PLN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

#### 5.4 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, conhecido como BERT, 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. Isso significa que ele considera as 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 prever estas 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 a 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 modelos como o BERT 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 a 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.



Figura 6 - Etapas do modelo BERT



Adaptado de HUGGING FACE (2024)

Na Figura 6 pode ser observado o fluxo de funcionamento do modelo BERT, onde em cada uma das etapas:

- **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 a "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, que é uma representação numérica do token. Os *embeddings* são semelhantes a 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. Este processo é semelhante à filtragem 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

## 5.5 Modelos de Linguagem Grande (LLMs)

Modelos de Linguagem Grande (LLMs) 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 lidamos 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 geração de documentos legais, refletindo a interseção entre a tecnologia de processamento de linguagem natural e a 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 a 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), que aponta em seu trabalho, testes com um modelo avançado de LLM que 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 a questões médicas complexas, abrindo caminhos para uma assistência ao paciente mais informada e eficaz. 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 não só representam um avanço tecnológico significativo no campo do aprendizado de máquina, mas também 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).

#### **5.5.1 O modelo gpt3.5-turbo**

Após examinar a evolução geral dos Modelos de Linguagem Grande (LLMs) na subseção anterior, nesta seção serão destacadas características de um modelo específico: o GPT-3.5-Turbo pela OpenAI. Este é um modelo avançado de linguagem natural desenvolvido

pela, representando a mais recente evolução nos LLMs (OpenAI, 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 com o 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, 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(OpenAI, 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. Sua 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.

Tabela 3 - Exemplo de prompt de chamada para o GPT3.5-turbo

Prompt	Resultado	Tempo de execução
<p>Dada a seguinte série de sentenças:</p> <p>'O paciente João da Silva tem 30 anos.'</p> <p>Analise as sentenças acima e identifique qual a classe FHIR mais provável em cada uma delas. Assuma a classe mais provável para cada sentença e, em seguida, crie uma visualização JSON, sem nenhum tipo de comentário adicional</p>	<pre>{   "resourceType": "Patient",   "name": [     {       "given": [         "João"       ],       "family": "Silva"     }   ],   "birthDate": "1994-01-07",   "age": {     "value": 30,     "unit": "anos",     "system": "http://unitsofmeasure.org"   } }</pre>	3 segundos
<p>Dada a seguinte série de sentenças:</p> <p>'paciente allan, 30 anos, tomou dipirona'</p> <p>Analise as sentenças acima e identifique qual a classe FHIR mais provável em cada uma delas. Assuma a classe mais provável para cada sentença e, em seguida, crie uma visualização JSON, sem nenhum tipo de comentário adicional</p>	<pre>{   "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"   } } ]</pre>	6 segundos

Fonte: Elaborado pelo autor (2023)

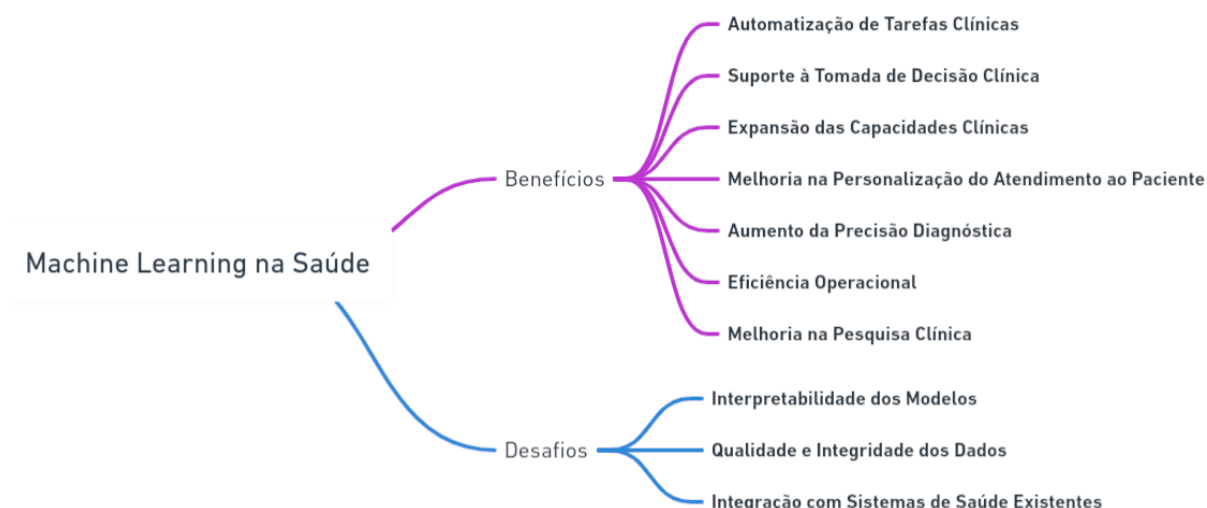
A tabela 3 exemplifica o uso do GPT3.5-turbo via prompt manual e sua performance de resposta medida em segundos. É necessário destacar que o mesmo tipo de uso pode ser feito com chamadas via API.

## 5.6 Desafios na Integração de ML e Saúde

No campo da Ciência da Informação, a combinação do *machine 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 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 a privacidade e a confidencialidade dos dados dos pacientes são primordiais, coletar e utilizar esses conjuntos de dados pode ser problemático. A preocupação com a ética da informação, 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, uma imagem que serve como uma ferramenta visual para consolidar e representar de forma estruturada os principais pontos abordados neste capítulo sobre "Machine Learning na Área da Saúde". Ao visualizar este diagrama, é possível rapidamente compreender as conexões entre os diferentes tópicos, facilitando a 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, onde entre os benefícios, se destaca a 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 também 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 também aumenta a eficiência, reduzindo custos e melhorando a gestão de recursos. No entanto, a figura também destaca os desafios associados ao uso do Machine Learning na saúde. Um dos principais é a interpretabilidade dos modelos, 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, e além disso, a integração de soluções de *Machine Learning* em sistemas de saúde existentes pode ser complexa, especialmente devido a 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 potenciais.

Tendo estabelecido a relevância e o potencial do ML na área da saúde, o próximo capítulo se aprofundará 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 fornecem contexto e significado, permitindo que os algoritmos de ML e os profissionais de saúde compreendam e utilizem os dados de forma mais eficaz e consigam validar os modelos. Assim, será explorado 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 inclusive podem ser fatais. A reação adversa a medicamentos, alimentos ou substâncias pode variar de sintomas leves a reações anafiláticas graves, o que torna muito importante a correta identificação e registro dessas alergias. No contexto do Hospital Sírio Libanês, um dos mais renomados centros médicos do país, a gestão e análise eficaz desses dados torna-se ainda mais relevante. Ao integrar metadados de negócio na análise dessas informações, é possível aprimorar a precisão, contextualização e aplicabilidade dos dados de alergia anonimizados, garantindo uma assistência médica mais informada e segura para cada paciente.

### **6.2 Importância de dados de alergia**

No contexto da saúde, a importância dos dados de alergia, tanto alimentares quanto a medicamentos, 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 a 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.

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) uma parcela significativa de pacientes investigados por suspeita de alergia a medicamentos, na realidade, não apresenta alergias verdadeiras. No estudo, 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 particularmente

relevante no contexto do tratamento com antibióticos, como as penicilinas, onde o diagnóstico preciso de alergias verdadeiras versus suspeitas não confirmadas permite a seleção de tratamentos mais eficazes e com menor risco de efeitos colaterais adversos.

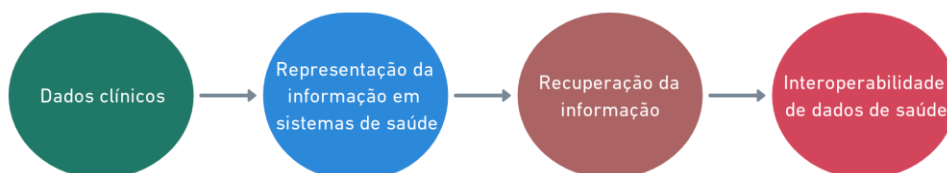
Portanto, a importância de abordagens detalhadas e cuidadosas no diagnóstico de alergias, seja alimentar ou a medicamentos na saúde é essencial. A gestão eficaz de alérgenos são fundamentais para reduzir riscos e melhorar a qualidade do cuidado ao paciente, e estas abordagens são essenciais para o desenvolvimento 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 destaca a importância da gestão eficiente de informações médicas. A próxima seção introduz a Arquitetura da Informação (AI) como uma ferramenta importante para esta finalidade, pois facilita o 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 a inadequada representação das informações nos sistemas da Câmara dos Deputados como um problema, é apresentada uma metodologia de Arquitetura da Informação proposta consistindo 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. Um dos artefatos apresentados é a matriz de metadados, que concentra as informações centrais necessárias para os processos de trabalho, sendo o principal entregável pois auxilia tanto no gerenciamento das informações, como no processo de desenvolvimento de sistemas 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 a informação de dados clínicos*



Fonte: Adaptado de Brandt(2020)

A Figura 8 exemplifica o fluxo de representação e recuperação da informação contida em prontuários eletrônicos. Baseando-se nela, 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 a avaliação de pacientes sob o 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 é "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, vemos "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, temos "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 PI", "NUTRICIONISTA CLINICO", "MEDICO".
- TIPO\_ATENDIMENTO: Esta coluna está repetida, já que já foi mencionada anteriormente na lista.

A seguir, a tabela uma amostra do *dataset* original:

Tabela 4 - Amostra original dos dados de alergia do HSL

#	NM ESTABELECIMENTO	DS SETOR	TIPO ATENDIMENTO	ADULTO PEDIATRIA	DT NASCIMENTO	DT REGISTRO	IE NEGA ALERGIAS	IE INTENSIDADE	AGENTE CAUSADOR	DS PRINCIPIO <sub>4</sub>	DS PRINCIPIO	CLASSE MATERIAL	DS OBSERVACAO	DS REACAO	FUNCAO	TIPO ATENDIMENTO
1	Hospital Sírío Libanês	Unidade Semi Intensiva (USI) - C - 06º andar	Internado	Adulto	11/27/1981	1/1/21 11:54	N	L	Nega Alergias						Enfermeiro PI	Internado
2	Hospital Sírío Libanês	Unidade Coronária - D - 07º andar - Ala III	Internado	Adulto	1/19/1969	1/1/21 11:56	N	L	Medicamentos	Antibacterianos, Penicilínicos de amplo espectro		Antibacterianos, Penicilínicos de amplo espectro		Erupção Cutânea	Enfermeiro PI	Internado
3	Hospital Sírío Libanês	Unidade Internação - D - 17º andar - Ala II	Internado	Adulto	7/24/1936	1/1/21 12:11	S	L		Nega alergia alimentar					NUTRICIONISTA CLINICO	Internado
4	Hospital Sírío Libanês	Unidade Internação - D - 18º andar - Ala II	Internado	Adulto	7/17/1929	1/1/21 12:12	S	L		Nega alergia alimentar			ajuste		NUTRICIONISTA CLINICO	Internado
5	Hospital Sírío Libanês	D - UAIC - Ala VI	Internado	Adulto	9/28/1961	1/1/21 12:13	N	L	Outro	piridoxina				Erupção Cutânea	MEDICO	Internado
6	Hospital Sírío Libanês	D - UAIC - Ala VI	Internado	Adulto	9/28/1961	1/1/21 12:13	N	L	Medicamentos		Metoclopramida			Edema de Glote	MEDICO	Internado
7	Hospital Sírío Libanês	D - UAIC - Ala VI	Internado	Adulto	9/28/1961	1/1/21 12:13	N	L	Medicamentos		Paracetamol			Erupção Cutânea	MEDICO	Internado

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, e então 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*

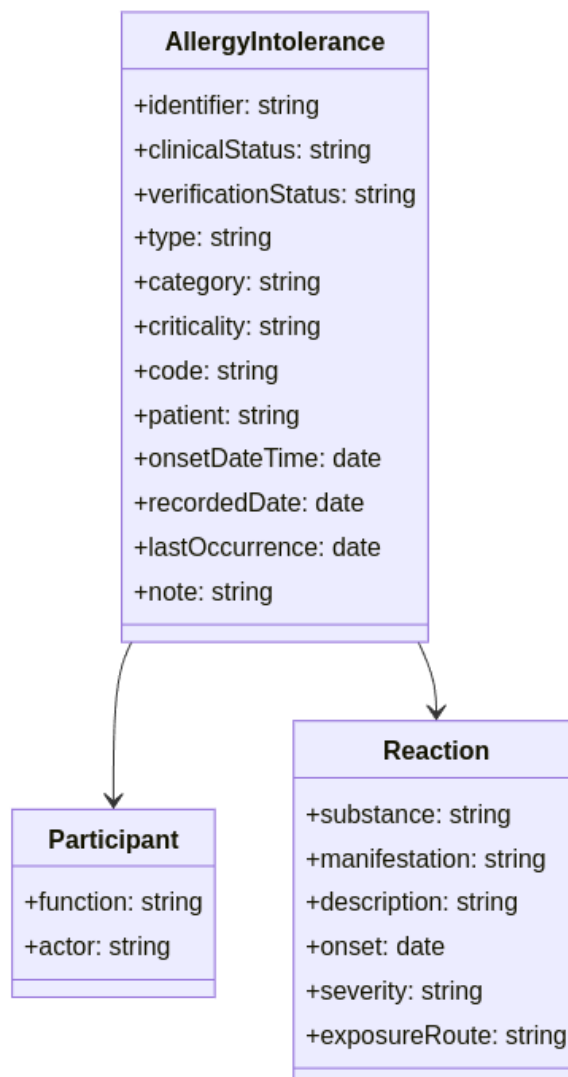


Fonte: Elaborado pelo autor

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. A classe *AllergyIntolerance* do padrão FHIR foi concebida com o propósito de representar informações relativas a 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* da classe *AllergyIntolerance*, que especifica a substância alergênica. A descrição da reação alérgica, representada no campo DS\_REACAO do conjunto de dados, teve sua correspondência na classe *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* da classe 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* da classe *AllergyIntolerance*.

A capacidade de rastrear quando uma alergia foi identificada e registrada mostrou-se crucial para a gestão clínica e para entender a evolução do paciente ao longo do tempo. Ao considerar todas essas correspondências diretas e a natureza intrínseca da classe *AllergyIntolerance* para representar alergias e intolerâncias, tornou-se evidente que esta era a classe FHIR mais adequada 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 a classe *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.
- lastoccurrence: a última ocorrência da reação.
- note: quaisquer observações ou anotações adicionais sobre a intolerância ou alergia.

Associado à classe *AllergyIntolerance*, há uma classe chamada *Participant*. Esta classe detalha quem registrou ou observou a alergia ou intolerância. 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 a classe *Reaction*, que descreve a reação específica que o paciente teve à substância alergênica. Os atributos desta classe incluem:

- substance: a substância específica que causou a reação.
- manifestation: a manifestação clínica da reação.
- description: uma descrição detalhada da reação.
- onset: o início da reação.
- severity: a gravidade da reação.
- exposureRoute: o 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 a classe 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 afim de identificar não só o contexto de produção de cada atributo mas também mas também do ponto de vista de negócios. No contexto empresarial, cada campo de informação pode ter implicações significativas para a 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, e 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 a 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 a questões de privacidade, segurança ou outras razões regulatórias.

Considerando a Matriz de Metadados proposta e o contexto do trabalho, ajustes foram realizados e então duas colunas adicionais foram adicionadas: a primeira para a 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 a Matriz de Metadados, disponibilizada na tabela a seguir:

Após a formatação desta tabela, obteve-se a Matriz de Metadados:

Tabela 5 - Metadados de negócio gerados com interoperabilidade com o padrão FHIR

ID	Nome	Contexto	Definição	Gestor	Gestor do Dado	Forma de acesso	Entrada padronizada	Regra formato	Alimentação inicial	Dados abertos	Restrição de acesso	Conceito FHIR	Vocabulário
1	Setor	Identificar a equipe de atendimento	Diferentes equipes que realizaram os procedimentos de avaliação com o objetivo de identificação de alergias.	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Unidade 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	Textual	ERP HSL	Não	Privado	Não se aplica	
2	Tipo Atendimento	Identificar a modalidade de atendimento	Diferentes modalidades para identificar as acomodações e agendamento do contato assistencial	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Pronto Atendimento   Internado	Textual	ERP HSL	Não	Privado	Não se aplica	
3	Adulto Pediatria	faixa etária do paciente	Indica o tipo de cuidado assistencial necessário mediante a idade	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Adulto   Pediatria	Textual	ERP HSL	Não	Privado	Não se aplica	
4	Data Nascimento	Identificar idade do paciente	Dia, mês e ano do nascimento do paciente	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Não se aplica	Data	ERP HSL	Não	Privado	Patient:birth Date	
5	Data Registro	Identificar o período do atendimento	Dia, mês e ano do nascimento do contato assistencial	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Não se aplica	Data	ERP HSL	Não	Privado	AllergyIntolerance:recordedDate	
6	Intensidade	Indica a severidade da ocorrência do sintoma	Diferentes níveis de manifestações alérgicas	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	I   D   L   M	Data	ERP HSL	Não	Privado	AllergyIntolerance:Manifestation:severity	fhir/ValueSet/reaction-event-severity
7	Agente Causador	Identificação do mecanismo fisiológico para um Risco de Reação	Diferentes origens causadoras do sintoma/ reação alérgica	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	food   medication   environment   biologic	Textual	ERP HSL	Não	Privado	AllergyIntolerance:category	fhir/ValueSet/allergy-intolerance-category
8	Princípio	Identificação do elemento causador das manifestações	Diferentes produtos, alimentos responsáveis pelo desencadeamento	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Não se aplica	Textual	ERP HSL	Não	Privado	AllergyIntolerance:substance	SNOMED
9	Reação	Identificar as reações ou manifestações que evidenciam a presença de alergia	Diferentes eventos de reação adversa ligados à exposição à substância	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Agitação psicomotora   Anafilaxia   Broncoespasmo   Choque anafilático   Coriza   Dermateite   Edema   Edema de glote   Erupção cutânea   Intolerância   Parada cardíaca   Prurido   Tosse	Textual	ERP HSL	Não	Privado	AllergyIntolerance:reaction:substance:coding:code	SNOMED
10	Função	Identificar quem ou o que participou das atividades relacionadas à alergia ou intolerância e como se envolveu	Diferentes funções dos profissionais de saúde	Coordenação ambulatorial	Coordenação ambulatorial	ERP HSL	Auxiliar Enfermagem   Assistente de enfermagem   Enfermeiro   Farmacêutico   Médico   Médico Residente   Nutricionista Clínico   Preceptor   Técnico Enfermagem	Textual	ERP HSL	Não	Privado	AllergyIntolerance:Participant:function	fhir/ValueSet/participation-role-type

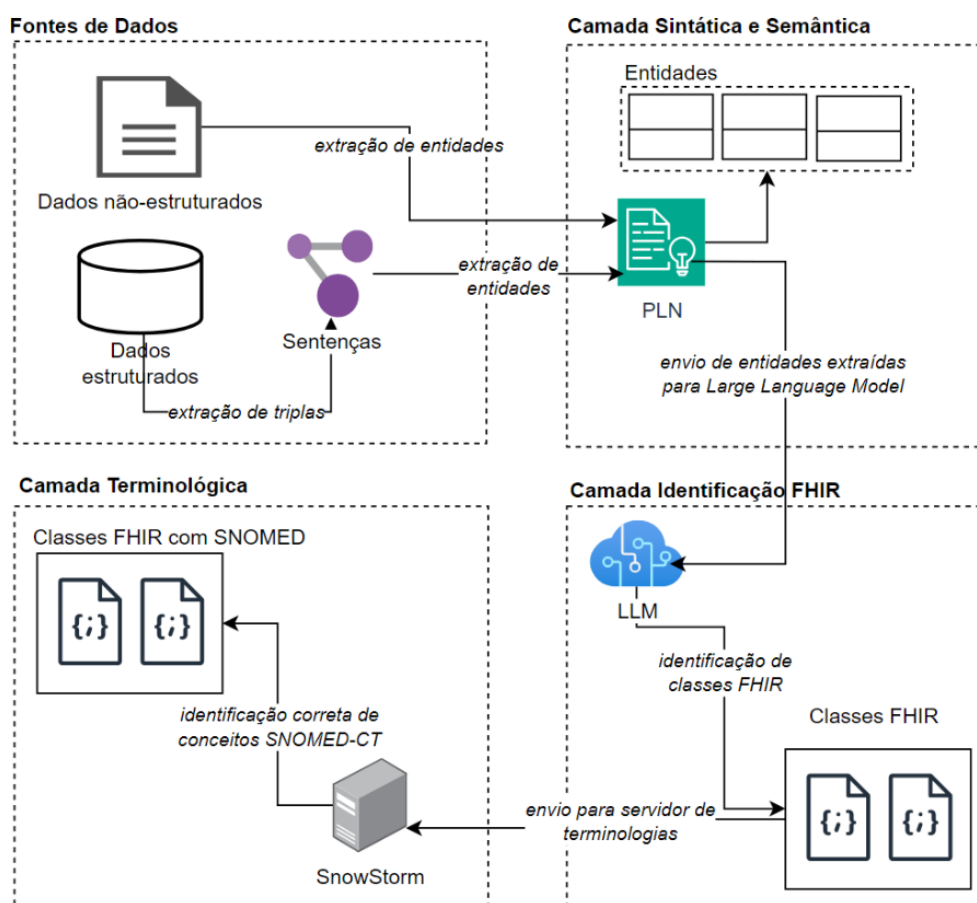
Após a conclusão da análise dos dados de alergia fornecidos pelo Hospital Sírio Libanês, ficou evidenciada a importância dos metadados de negócio no contexto dos sistemas de saúde. Essa análise demonstrou que, sem uma estruturação e interoperabilidade adequadas dos dados, a qualidade e aplicabilidade 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 no capítulo subsequente, será destinado à criação de um modelo informacional para a interoperabilidade de dados de saúde. Nessa fase, o *Machine Learning* será empregado para mapear dados conforme o padrão FHIR, permitindo uma abordagem mais avançada e precisa para a integração dos sistemas de saúde.

## **7 FHIRFLOW: MODELO DE MAPEAMENTO DE DADOS DE SAÚDE PARA FHIR**

Este capítulo introduz um modelo de interoperabilidade de dados de saúde, especificamente projetado para o 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 (PLN) e Modelos de Linguagem Grande (LLM), como o GPT-4, ele é capaz de analisar, estruturar e mapear dados de saúde com alta precisão. Este capítulo 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 (2023)

A Figura 11 representa uma visão geral do FHIR-FLOW, que em sua essência, representa um mecanismo abrangente que atua como uma ponte, transformando dados brutos e heterogêneos em informações padronizadas e interoperáveis. Em sua operação, o modelo assume a responsabilidade de processar, interpretar e mapear dados de saúde, seja eles estruturados ou não. Ele começa com uma análise sintática, normalizando e estruturando os dados em formas que serão posteriormente processadas com técnicas avançadas de PLN. O uso de modelos de linguagem, como os LLM, permite que o modelo compreenda e gere textos com relevância e coerência, destacando entidades e suas respectivas relações. À medida que o modelo avança, ele integra camadas de identificação semântica e terminológica focando nas classes FHIR e reconhecendo variações em vocabulários como SNOMED-CT, garantindo a consistência, alinhamento e interoperabilidade das informações.

Na primeira etapa, a 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 a 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, a análise semântica, é introduzida uma camada de identificação de classes FHIR, essencial para a 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 focando na 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 classes FHIR geradas a partir de dados, sejam eles estruturados ou não estruturados, permitindo uma comunicação integrada e harmonizada entre sistemas de informação em saúde distintos.



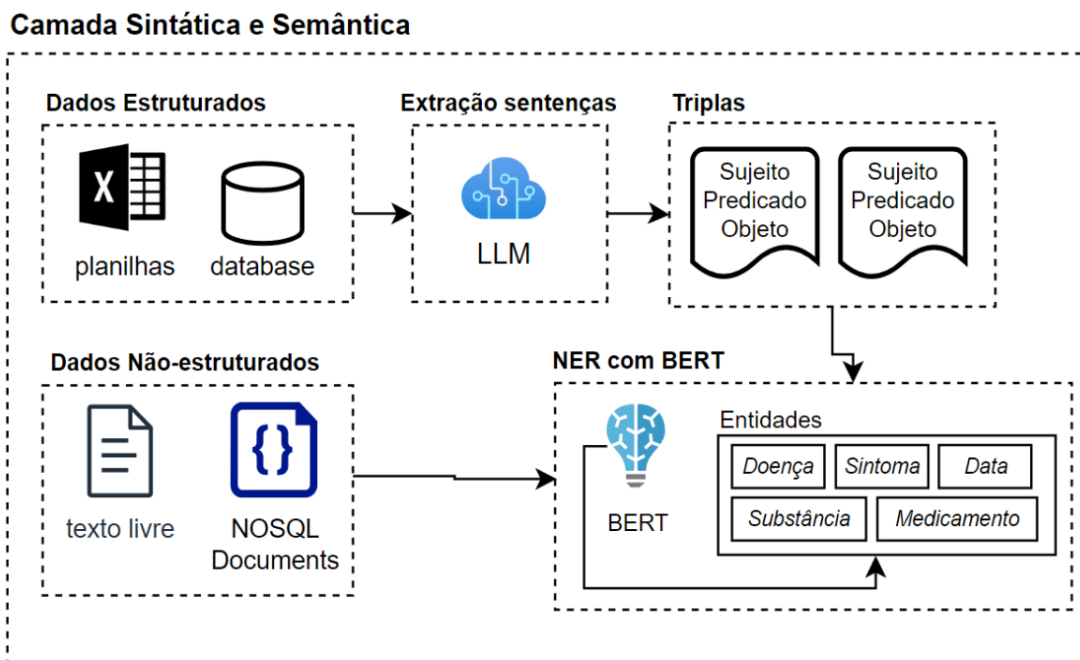
## 7.1 Camada Sintática e Semântica

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. Aqui, a ênfase é colocada na análise e estruturação dos registros de saúde, tanto estruturados quanto não estruturados, para facilitar sua posterior interpretação e integração. Essencialmente, a camada sintática e semântica serve como o ponto de entrada para os dados no sistema, onde são submetidos a um processo de normalização e análise gramatical.

O objetivo principal desta camada é transformar os registros eletrônicos de saúde em formatos que são tanto coerentes para a interpretação humana quanto compatíveis com processamentos automáticos posteriores e isso 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. 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 as entidades identificadas, que por sua vez são requisitos para o mapeamento de dados para classes 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 (2023)

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

- **Entrada de EHR:** Os registros eletrônicos de saúde (EHR) 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 os 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 abaixo:

Tabela 6 - Exemplo de extração de triplas a partir de fontes de dados

Dado Original	Triplas																									
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 autoinjjetável para possíveis casos de anafilaxia futura.	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 autoinjjetável 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																									
<table><tr><th>ID</th><th>Data</th><th>CID</th><th>Substância</th><th>Sintoma</th></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>Pasta de amendoim</td><td>Inchaço labial</td></tr></table>	ID	Data	CID	Substância	Sintoma	1335	2021-09-10	F1001	Pasta de amendoim	Desconforto	1335	2021-09-10	F1001	Pasta de amendoim	Prurido	1335	2021-09-10	F1001	Pasta de amendoim	Sinais de Urticária	1335	2021-09-10	F1001	Pasta de amendoim	Inchaço labial	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
ID	Data	CID	Substância	Sintoma																						
1335	2021-09-10	F1001	Pasta de amendoim	Desconforto																						
1335	2021-09-10	F1001	Pasta de amendoim	Prurido																						
1335	2021-09-10	F1001	Pasta de amendoim	Sinais de Urticária																						
1335	2021-09-10	F1001	Pasta de amendoim	Inchaço labial																						

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 ciência da informação e as técnicas de PLN, demonstrando a importância de abordagens estruturadas na interpretação de dados não estruturados.

Já 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 a 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 NER (*Named Entity Recognition*), é uma técnica importante que identifica e classifica entidades em textos, tais como pessoas, organizações, locais e datas e é essencial para a 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.

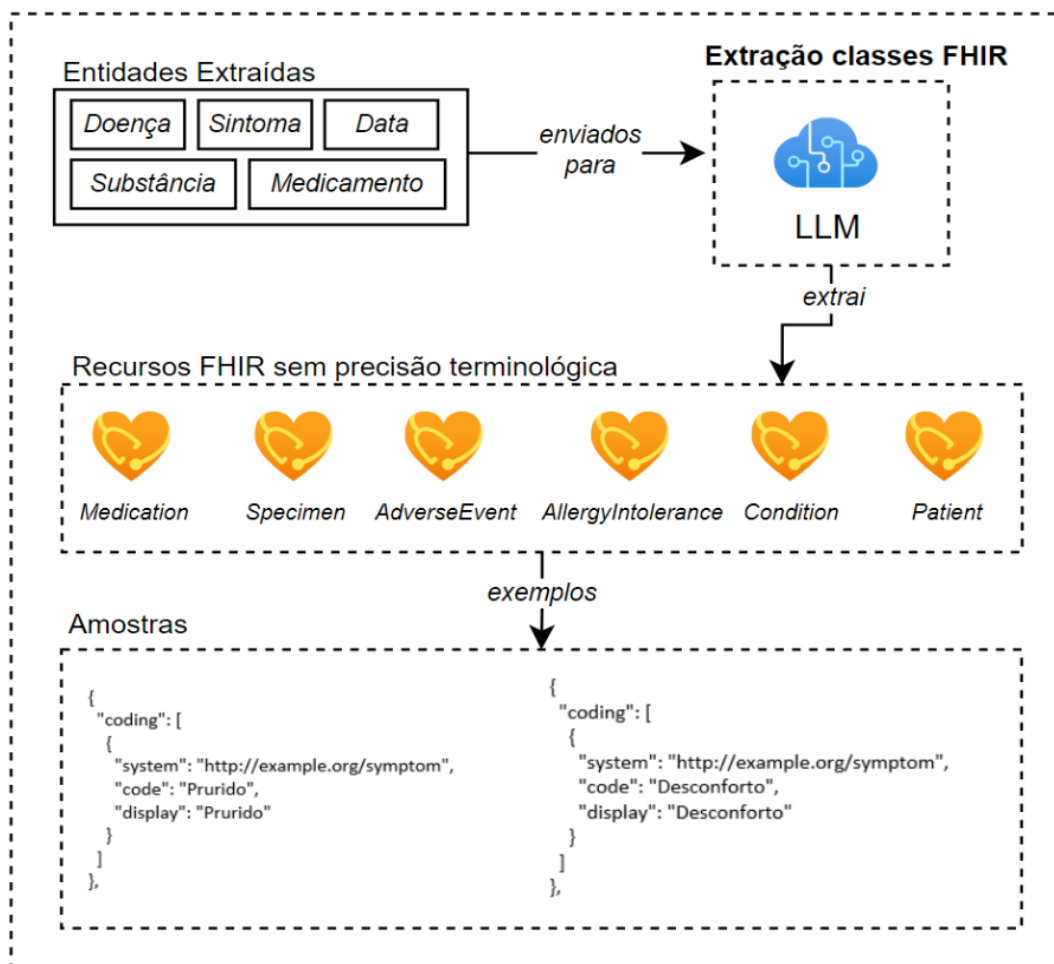
No contexto da saúde, a identificação de entidades nomeadas é especialmente relevante, pois pode impactar diretamente na qualidade e eficácia do atendimento ao paciente, pois como destaca Yang et al. (2022), os dados de texto médico registram dados clínicos detalhados e o reconhecimento de entidades nomeadas é a base do processamento de informações textuais e 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 classes 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, um sistema de NER pode reconhecer "hipertensão" como uma *Condition* (Condição) e o nome do medicamento como uma *Medication* (Medicação), atribuindo, consequentemente, estas informações às classes 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 as classes do padrão FHIR. As classes FHIR são essenciais para a organização de dados em categorias específicas, como Recursos, Extensões, Perfis e Operações. Nesta etapa do FHIR-FLOW, é utilizado o GPT-4 da OpenAI, 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 etapa, pois seu conhecimento semântico e contextual avançado é aplicado para predizer, com alta precisão, as classes FHIR mais adequadas para cada tripla, porém sem a responsabilidade de precisão terminológica.

Figura 10 - Camada de identificação FHIR

**Camada Identificação FHIR**

Fonte: Elaborado pelo autor (2023)

Conforme ilustra a Figura 10, nesta etapa foi utilizado um modelo de LLM, o GPT-4 da OpenAI (OpenAI, 2023), utilizado 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 as possíveis classes FHIR correspondentes. O conhecimento semântico e contextual do modelo de linguagem permite a associação das triplas às classes FHIR mais prováveis.

A seguir, uma tabela exemplificando uma entrada e saída desta etapa de identificação

FHIR:

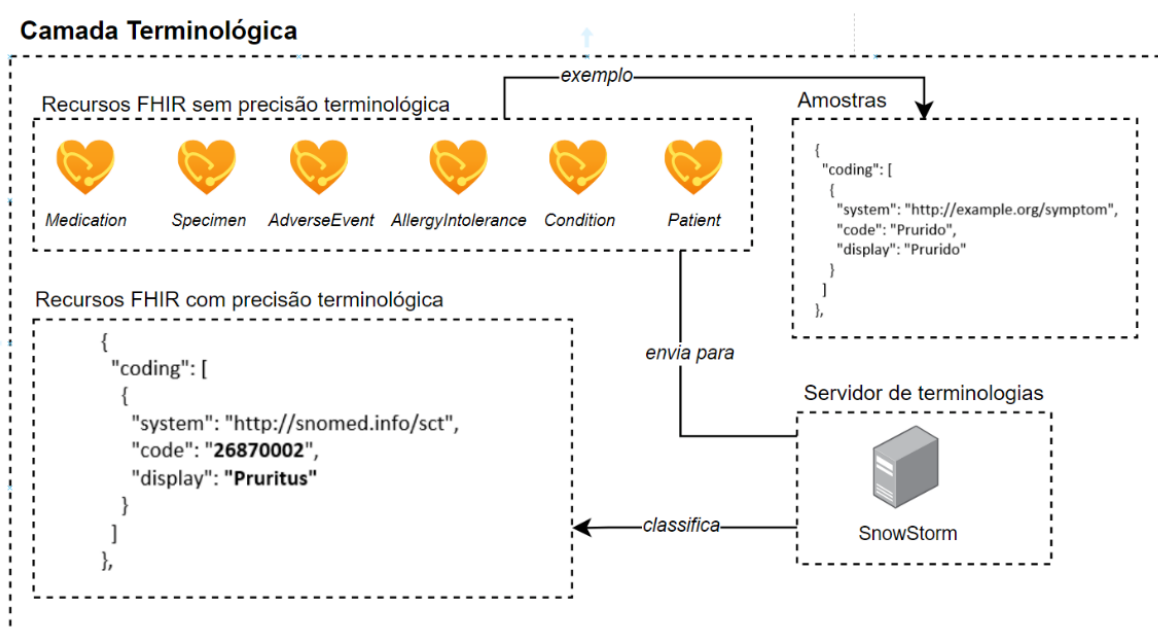
*Tabela 7 - Exemplo de identificação de classes FHIR através de triplas*

Triplas	Classe 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	<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": "InchacoLabial",           "display": "Inchaço Labial"         }       ]     }   ] } </pre>

### 7.3 Camada Terminológica

Esta subseção explica o processo para a identificação terminológica dos códigos encontrados na etapa anterior, onde as classes FHIR foram identificadas. Para esta fase, o modelo se restringirá às terminologias da SNOMED, utilizando o SNOWSTORM como servidor de terminologias.

Figura 11 - Camada Terminológica



Fonte: Elaborado pelo autor (2023)

Como mostra a Figura 11, o processo de identificação terminológica é baseado em requisições feitas via API ao servidor SNOWSTORM, com o objetivo de transformar as classes extraídas na etapa anterior em classes com os códigos corretos da SNOMED. Cada classe identificada é submetida a uma consulta ao servidor SNOWSTORM, onde os termos correspondentes na SNOMED são identificados e vinculados à classe. Esta etapa é importante para garantir que cada termo e conceito utilizado seja alinhado com a terminologia padrão da SNOMED, promovendo consistência e precisão na interoperabilidade.

A tabela 8 exemplifica uma classe FHIR extraída na camada anterior (identificação FHIR) depois de passar pela camada de identificação terminológica:



Tabela 8 - Exemplo de identificação terminológica em classe FHIR

Classe FHIR sem terminologia	Classe 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 (2023)

### 7.3.1 A SNOMED-CT

A SNOMED-CT é uma terminologia clínica internacionalmente reconhecida, sendo um recurso preciso e abrangente para a representação sistemática de termos clínicos. 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.

Entretanto, a adoção da SNOMED-CT 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 o 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, uma tabela demonstrando aplicações de uso da SNOMED em diversas áreas:

Tabela 9 - Uso da SNOMED-CT em diversas áreas da saúde

Área de Aplicação	Descrição	Benefícios
<b>EMR</b>	Utilização 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.
<b>Pesquisa Clínica</b>	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.
<b>Gestão em Saúde</b>	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.
<b>Farmacovigilância</b>	Emprego na identificação e análise de eventos adversos relacionados a medicamentos.	Contribui para a segurança do paciente e monitoramento eficaz de medicamentos no mercado.
<b>Interoperabilidade</b>	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.
<b>Apoio à Decisão Clínica</b>	Utilização 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.
<b>Codificação e Classificação</b>	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 (2024).

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 O 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, e como servidor de terminologia focado no SNOMED-CT, o Snowstorm implementa funcionalidades específicas para esta nomenclatura, incluindo o URI (*Uniform Resource Identifier*) do SNOMED CT, filtros, propriedades, conjuntos de valores e mapas de conceitos implícitos (IHTSDO,2023).

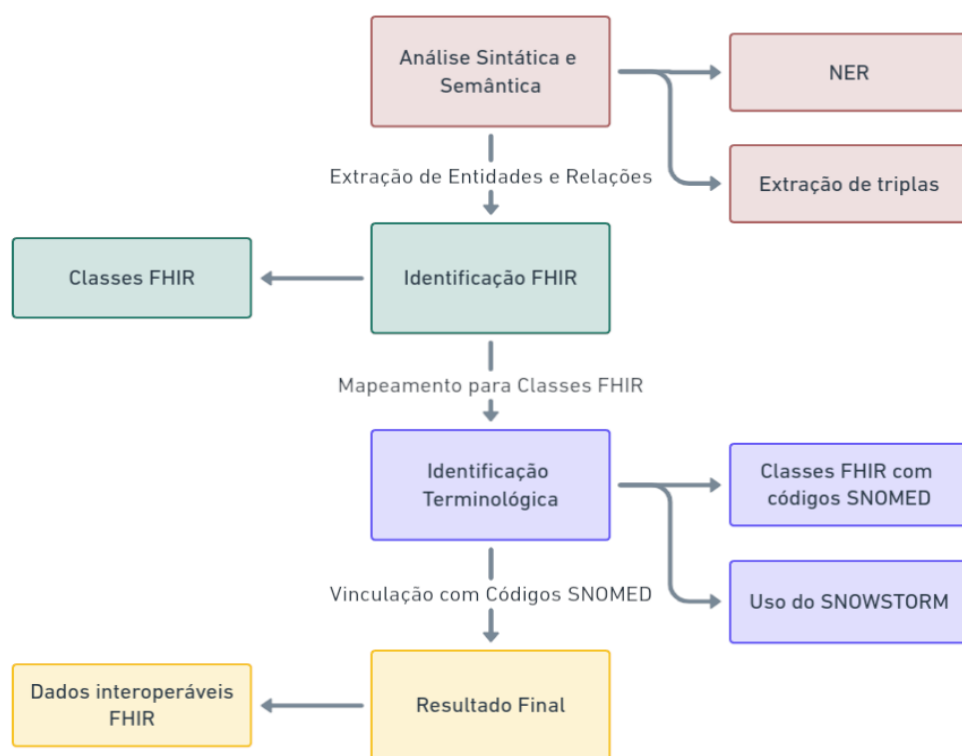
O suporte para o Snowstorm é fornecido pela comunidade através do repositório do GitHub, já que o SNOMED International não oferece suporte comercial. O produto possui duas APIs principais: a FHIR API, que implementa o módulo de terminologia e suporta diversos sistemas de código, e a API Especialista 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 a recursos terminológicos avançados, como a 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 classes FHIR, vinculação de códigos SNOMED-CT, e classes FHIR finais devidamente classificadas.

Figura 12 - Visão geral do modelo



Fonte: Elaborado pelo autor (2023)

A Figura 12 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, então, levadas à segunda etapa, "Identificação FHIR", onde são mapeadas para classes FHIR correspondentes através de modelos de linguagem avançados como o GPT-4. Na terceira e última etapa, "Identificação Terminológica", as classes 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 e 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. Aqui estão os principais objetivos da validação:

1. 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.

2. 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 estão construídas representando as classes 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	<pre> "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"   } }, </pre>	<pre> "resourceType": "Condition",   "clinicalStatus": {     "coding": [       { "system": " http://snomed.info/sct", "code": 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) "         } ] } ] ] } </pre>
Anamnese	Mulher de 30 anos apresenta queixas de dores abdominais recorrentes, especialmente após as refeições. Relata episódios de diarreia e constipação alternados.	<pre> "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"     }   }, </pre>	<pre> {   "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"       }     ] ] } </pre>

Fonte: Elaborado pelo Autor (2023)

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 as classes 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	<pre> "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"     }]   }, </pre>	<pre> {   "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"       } ] } } </pre>
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.	<pre> {   "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"     }]   }, </pre>	<pre> "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"         } } } </pre>

Fonte: Elaborado pelo Autor (2023)

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 classes FHIR corretas, como *Patient* para detalhes do paciente e *Condition* para os sintomas relatados. Isto demonstra necessidade de ajuste no modelo, mas também a eficiência do modelo 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 as classes 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 os 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.	<pre> "resourceType": "Patient", "identifier": [{ "use": "usual", "type": { "coding": [{ "system": " http://snomed.info/sct ", "code" : "MR" }] }, </pre>	<pre> "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 cashews." }], "severity" : "severe" }] } </pre>
História de alergia	<p>Paciente: 35 anos</p> <p>Data de Identificação da Alergia: 10/03/2024</p> <p>Substância Alergênica: Amendoim</p> <p>Tipo de Reação: Alergia: Urticária</p> <p>Edema de Glote</p> <p>Dificuldade Respiratória</p> <p>Gravidade: Grave</p> <p>Último Episódio: 15/03/2024</p> <p>Informações Adicionais: O paciente teve sua primeira reação alérgica a amendoins aos 30 anos, caracterizada por urticária e dificuldade respiratória</p>	<pre> "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" } ] } </pre>	<pre> "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" } ] }, </pre>

Fonte: Elaborado pelo Autor (2023)

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 a classe 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, de 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 da classe *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, observamos 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 as classes FHIR corretas, 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, mas 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 uma classe 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 classes e atributos como *Patient*, *Condition*, e *AllergyIntolerance*, bem como 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

O modelo conseguiu identificar e gerar o recurso FHIR correspondente para 23 registros classificando os recursos como *AllergyIntolerance*, totalizando um percentual de classificação de 40% considerando o total geral de registros, mas 90% considerando os registros que tinham todas as informações completas de reação alérgica.

Figura 13 - Amostra de validação de texto livre com a aplicação rodando

→ 127.0.0.1:5000

## Modelo de interoperabilidade de dados de saúde utilizando FHIR

Texto Livre [Dados Estruturados](#)

Digite o texto:

Paciente: 35 anos

Data de Identificação da Alergia: 10/03/2024

Substância Alergênica: Amendoim

Tipo de Reação Alergica:

Urticária

Edema de Glote

Dificuldade Respiratória

Gravidade: Grave

Último Episódio: 15/03/2024

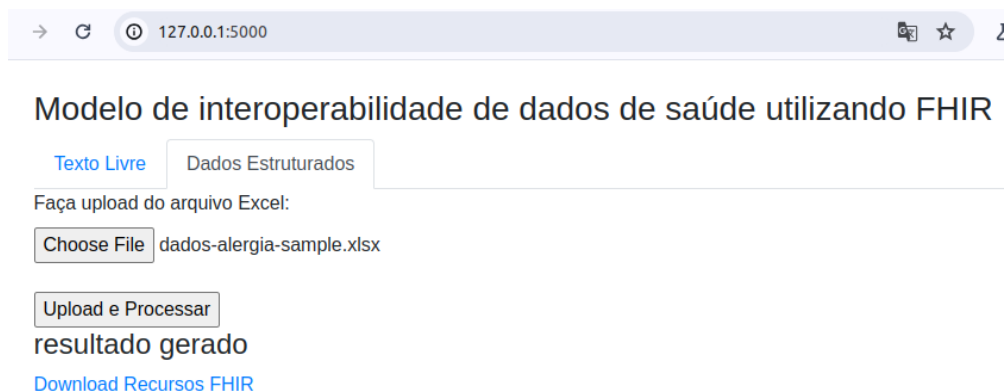
Processar

Fonte: Elaborado pelo Autor(2023)

A Figura 13 exemplifica o processo de validação dos dados com a tela da aplicação do modelo implementado, onde 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 14 - Tela de validação de dados estruturados e download do resultado*



Fonte: Elaborado pelo Autor(2023)

A Figura 14 mostra a tela para inserção de arquivo e validação de dados estruturados, bem como a opção para realizado o download com os 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 um modelo arquitetônico e a 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 seus capítulos 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 nos capítulos finais. 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 demonstram o impacto positivo de um modelo arquitetural para interoperabilidade aprimorado com uso de tecnologias como *Machine Learning* e LLMs associados a 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 pesquisa enfocou a importância da interoperabilidade em sistemas de saúde e o papel do padrão FHIR como um facilitador neste contexto. Foi observado que a aplicação de um modelo arquitetônico adequado, baseado no padrão FHIR, oferece possibilidades para melhorar a gestão e o uso de dados de saúde. Isso não apenas promove a eficiência operacional, mas também melhora a qualidade do atendimento ao paciente, ao permitir o acesso e a integração de dados de saúde de forma mais coesa e abrangente.

Além disso, esta pesquisa destacou a importância da interdisciplinaridade, unindo aspectos da ciência da informação, ciência da computação e cuidados de saúde, para abordar os desafios de interoperabilidade. A abordagem adotada revelou que soluções inovadoras podem surgir da interseção dessas disciplinas.

## 8.2 Contribuições

A pesquisa contribui significativamente para a compreensão da interoperabilidade de dados de saúde e do papel do padrão FHIR. As contribuições específicas incluem:

- O desenvolvimento de um modelo de arquitetura para a interoperabilidade de dados de saúde, baseado no padrão FHIR.
- Uma análise aprofundada do estado atual da interoperabilidade em saúde e dos desafios associados.
- Propostas de soluções e recomendações para a adoção efetiva do padrão FHIR em sistemas de saúde.
- Um modelo demonstrando a aplicabilidade prática do modelo proposto, oferecendo *insights* valiosos para futuras implementações.

Em conclusão, este estudo ressalta a importância da interoperabilidade em sistemas de saúde e a necessidade de métodos inovadores para melhorar a integração e gestão de dados de saúde. A adoção do padrão FHIR é destacada como um passo fundamental, transformando a maneira como os dados de saúde são compartilhados e gerenciados. O FHIR facilita a troca de informações entre diversas plataformas de saúde. A pesquisa também enfatiza os desafios na implementação do FHIR, incluindo a necessidade de infraestruturas tecnológicas sólidas e políticas de governança de dados eficientes.

Em resumo, o estudo demonstra que o FHIR é um divisor de águas na gestão de dados de saúde, prometendo um futuro com informações de saúde mais acessíveis e úteis, e que por isso, soluções para habilitar a interoperabilidade para o seu padrão, como o FHIR-FLOW, tem impacto direto na qualidade de atendimento do paciente.



## REFERÊNCIAS

- BRANDT, Mariana Baptista. **Modelagem da informação legislativa: arquitetura da informação para o processo legislativo brasileiro**. 2020.
- CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2003.
- DRESCH, A.; LACERDA, D. P.; JÚNIOR, J. A. V. A. **Design Science Research: método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2015
- Graber ML, Byrne C, Johnston D. **The impact of electronic health records on diagnosis. Diagnosis (Berl)**. 2017 Nov 27;4(4):211-223. doi: 10.1515/dx-2017-0012. PMID: 29536944.
- HÜNER, K. M.; OTTO, B; ÖSTERLE, H. **Collaborative management of business metadata**. International Journal of Information Management, v. 31, 2011, p. 366-373.
- INMON, W. H., O; NEIL, B. ; FRYMAN, L. **Business metadata: Capturing enterprise knowledge**. Morgan Kaufmann: Boston, 2008.
- MARCO, D. **Managing Metadata for the Business**, Part 1. DM Review, New York, v. 16, n. 2, p. 42-43, fev. 2006.
- MAZUCATO, Thiago et al. **Metodologia da pesquisa e do trabalho científico**. Penápolis: Funepe, 2018.
- NARDON, Fabiane Bizinella; MOURA JUNIOR, Lincoln de Assis. **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**. J Digit Imaging 32, 354–361 (2019). <https://doi.org/10.1007/s10278-018-0134-3>
- THIRU, Krish; HASSEY, Alan; SULLIVAN, Frank. **Systematic review of scope and quality of electronic patient record data in primary care**. Bmj, v. 326, n. 7398, p. 1070, 2003.
- Tierney MJ, Pageler NM, Kahana M, Pantaleoni JL, Longhurst CA. **Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions**. Acad Med. 2013 Jun;88(6):748-52. doi: 10.1097/ACM.0b013e3182905ceb. PMID: 23619078.
- PETRY, Karine et al. **Utilização do Padrão HL7 para Interoperabilidade em Sistemas Legados na Área de Saúde**. In: XI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE. 2008.
- RODRIGUES, M. R.; CERVANTES, B. M. N. Organização e representação do conhecimento por meio de mapas conceituais. **Ciência da Informação**, Brasília, v.43, n.1, p.154-169, jan./abr.2014. Disponível em: <http://revista.ibict.br/ciinf/article/view/1425/1603>
- SHERMAN, R. **Align Metadata and Business Initiatives**. DM Review, New York, v. 16, n. 1, p. 50, Jan. 2006.
- STAN, Ovidiu; MICLEA, Liviu. **Local EHR management based on FHIR**. In: 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). IEEE, 2018. p. 1-5

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

YAZAN, Bedrettin et al. **Três abordagens do método de estudo de caso em educação:**

MEHTA, Neil; DEVARAKONDA, Murthy V. **Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey?**. Journal of Allergy and Clinical Immunology, v. 141, n. 6, p. 2019-2021. e1, 2018.

JAVOID, Mohd et al. **Significance of machine learning in healthcare: Features, pillars and applications**. International Journal of Intelligent Networks, v. 3, p. 58-73, 2022.

BOUH, Mohamed Mehfood; HOSSAIN, Forhad; AHMED, Ashir. **A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format**. 2023.

BI, Qifang et al. **What is machine learning? A primer for the epidemiologist**. American journal of epidemiology, v. 188, n. 12, p. 2222-2239, 2019.

AYAZ, Muhammad et al. **The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities**. JMIR medical informatics, v. 9, n. 7, p. e21929, 2021.

SARIPALLE, Rishi; RUNYAN, Christopher; RUSSELL, Mitchell. **Using HL7 FHIR to achieve interoperability in patient health record**. Journal of biomedical informatics, v. 94, p. 103188, 2019.

HL7 FHIR. (2021). **AllergyIntolerance resource**. Disponível em: <https://www.hl7.org/fhir/allergyintolerance.html>.

PATEL, Ankur A; ARASANIPALAI, Ajay Uppili; **Applied Natural Language Processing in the Enterprise**; editora O'Reilly, 2021.

Caseli, H., Freitas, C., & Viola, R. (2022). **Processamento de Linguagem Natural**. Short courses of the 37th Brazilian Symposium on Data Bases, Búzios, RJ, Brazil.

Oliveira, B.S.N., Rêgo, L.G.C. do, Peres, L., Silva, T.L.C. da, & Macêdo, J.A.F. de. (2022). **Processamento de Linguagem Natural via Aprendizagem Profunda**. In: ALMEIDA, E. S.; SANTORO, F. M. (org.). 41ª Jornada de Atualização em Informática. Porto Alegre: SBC.

JAGANNATHA, Abhyuday et al. **Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0)**. Drug safety, v. 42, p. 99-111, 2019.

CASELI, Helena; FREITAS, Cláudia; VIOLA, Roberta. **Processamento de Linguagem Natural**. Sociedade Brasileira de Computação, 2022.

PADOVANI, Djalma. **Um método adaptativo para análise sintática do Português Brasileiro**. 2022. Tese de Doutorado. Universidade de São Paulo.

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

Yang, T., He, Y., & Yang, N. (2022). **Named Entity Recognition of Medical Text Based on the Deep Neural Network**. Journal of Healthcare Engineering, Volume 2022.

CORNET, R.; DE KEIZER, N. **Forty years of SNOMED: a literature review**. BMC Medical Informatics and Decision Making, v. 8, Suppl 1:S2, 27 out. 2008.

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

CHATTERJEE, Ayan; PAHARI, Nibedita; PRINZ, Andreas. 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.

PIMENTA, Nuno et al. Interoperability of Clinical Data through FHIR: A review. **Procedia Computer Science**, v. 220, p. 856-861, 2023.

BALCH, Jeremy A. et al. Machine Learning–Enabled Clinical Information Systems Using Fast Healthcare Interoperability Resources Data Standards: Scoping Review. **JMIR Medical Informatics**, v. 11, p. e48297, 2023.

ROEHRS, Alex et al. Toward a model for personal health record interoperability. **IEEE journal of biomedical and health informatics**, v. 23, n. 2, p. 867-873, 2018.

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

HUGGING FACE. **Documentação do modelo BERT**. Disponível em: <[https://huggingface.co/docs/transformers/model\\_doc/bert](https://huggingface.co/docs/transformers/model_doc/bert)>. Acesso em: 4 de janeiro de 2024.

IQBAL, Ashraf Mohammed; SHEPHERD, Michael; ABIDI, Syed Sibte Raza. An ontology-based electronic medical record for chronic disease management. In: **2011 44th Hawaii International Conference on System Sciences**. IEEE, 2011. p. 1-10.

JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685-695, 2021.

GHASSEMI, Marzyeh et al. A review of challenges and opportunities in machine learning for health. **AMIA Summits on Translational Science Proceedings**, v. 2020, p. 191, 2020.

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

Organização Internacional de Terminologia em Saúde (IHTSDO). (2023). **Snowstorm**. Disponível em: <https://github.com/IHTSDO/snowstorm>. Acesso em: dezembro de 2023.

FRANSSON, Sara et al. **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.

DE LA CRUZ, Silvia et al. Alergias alimentarias: **Importancia del control de alérgenos en alimentos**. Nutr Clín Diet Hosp, v. 38, n. 1, p. 142-148, 2018.

- OpenAI. **Visão geral dos modelos**. Disponível em:  
<https://platform.openai.com/docs/models/overview>. Acesso em: dezembro de 2023.
- KADDOUR, Jean et al. **Challenges and applications of large language models**. arXiv preprint arXiv:2307.10169, 2023.
- SINGHAL, Karan et al. **Large language models encode clinical knowledge**. Nature, v. 620, n. 7972, p. 172-180, 2023.