

CAPÍTULO 3

INTEGRAÇÃO ENTRE DADOS TEXTUAIS DE PRONTUÁRIOS ELETRÔNICOS DO PACIENTE (PEPS) E TERMINOLOGIAS CLÍNICAS

Data de aceite: 01/11/2021

Data de submissão: data 09/08/2021

Amanda Damasceno de Souza

Escola de Ciência da Informação, Universidade Federal de Minas Gerais
Belo Horizonte - Minas Gerais
<https://orcid.org/0000-0001-6859-4333>

Eduardo Ribeiro Felipe

Universidade Federal dos Vales do Jequitinhonha e Mucuri
Diamantina – Minas Gerais
<https://orcid.org/0000-0003-1690-2044>

Fernanda Farinelli

Instituto de Gestão em Tecnologia da Informação, Engenharia e Análise de Dados
Pós-graduação em Estudos Linguísticos,
Universidade Federal de Minas Gerais
Belo Horizonte - Minas Gerais
<https://orcid.org/0000-0003-2338-8872>

Jeanne Louize Emygdio

Escola de Ciência da Informação, Universidade Federal de Minas Gerais
Belo Horizonte - Minas Gerais
<https://orcid.org/0000-0002-7329-4447>

Lívia Marangon Duffles Teixeira

Instituto de Gestão em Tecnologia da Informação, FIEMG
Belo Horizonte - Minas Gerais
<https://orcid.org/0000-0001-9728-3905>

Maurício Barcellos Almeida

Escola de Ciência da Informação, Universidade Federal de Minas Gerais
Belo Horizonte - Minas Gerais
<https://orcid.org/0000-0002-4711-270X>

RESUMO: Um problema das Ciências de Vida, como um todo e em todo o mundo, é lidar com o grande volume de informação produzida constantemente em cada um dos inúmeros campos de pesquisa. Para isso, os profissionais da informação – analistas, cientistas da informação, cientistas da computação, bibliotecários, linguistas, dentro outros – se valem de terminologias médicas padronizadas. Entretanto, tais terminologias mantêm diversos problemas conhecidos que impactam na capacidade de sistemas de informação inteligentes em lidar com o grande volume de dados de forma automática. Para isso, apresentam-se as ontologias como alternativa de sucesso. O objetivo deste estudo é descrever aspectos necessários para o entendimento do problema, além de fornecer alternativas para integrar dados clínicos em PEP com terminologias clínicas padronizadas no formato de uma metodologia. A metodologia apresentada aqui como resultado reflete a investigação realizada em um hospital brasileiro de grande porte, onde analisaram-se dados de milhares de registros médicos reais, confrontando-os com a classificação que receberiam em bem conhecidas terminologias clínicas.

PALAVRAS-CHAVE: Terminologias Clínicas. Biblioteconomia Clínica. Interoperabilidade. Ontologias.

INTEGRATION BETWEEN TEXTUAL DATA FROM ELECTRONIC MEDICAL RECORDS (EHRs) AND CLINICAL TERMINOLOGIES

ABSTRACT: A problem for the Life Sciences, as a whole and throughout the world, is to address the large volume of information produced in the numerous research fields. To this end, information professionals – analysts, information scientists, computer scientists, librarians, linguists, among others – use standardized medical terminologies. However, such terminologies maintain several well-known problems impacting the ability of intelligent information systems to handle large volumes of data automatically. Ontologies are presented as a successful alternative for these issues. This study aims to describe aspects required to understand the problems and provide alternatives for the integration, on the one hand, clinical data in EHRs and, on the other hand, the standardized clinical terminology. The result of this process is presented in the form of a methodology. Such methodology presented here, as a result, reflects the investigation carried out in a large Brazilian hospital, where data from thousands of actual medical records were analyzed, comparing them with the classification they would receive in well-known clinical terminology.

KEYWORDS: Clinical Terminology. Clinical Librarianship. Interoperability. Ontology.

1 | INTRODUÇÃO

Um dos desafios da Medicina é lidar com o volume de informação produzida diariamente, além de gerenciar as diversas fontes e formatos. Para gerenciar os recursos informacionais, as instituições de saúde adotam diversos sistemas de informação (SI), heterogêneos e distribuídos, mas que necessitam trocar informações entre si.

Ao pensar na informação relevante para a prática médica, o prontuário médico é um dos documentos mais importantes. Trata-se do registro do atendimento do paciente e funciona como meio de comunicação para a equipe responsável pelos cuidados de saúde (MASSAD *et al.*, 2003). Tendo em vista a evolução das tecnologias de informação e comunicação (TICs), este documento assume formato eletrônico, passando a ser conhecido como “prontuário eletrônico do paciente” (PEP). Torna-se um repositório eletrônico de informação sobre estado de saúde e os cuidados ao indivíduo ao longo de toda sua vida (REIS; CORREIA; PEREIRA, 2011).

O PEP é concebido por equipes multidisciplinares que adotam tanto a linguagem natural quanto terminologias padronizadas para preenchê-lo. Tais terminologias são heterogêneas: têm um longo histórico, com origens no século XIX, e foram criadas por diversos motivos. No século XXI, as terminologias continuam essenciais para a padronização do vocabulário médico em sistemas computacionais. Tais sistemas revolucionaram o atendimento médico, mas são ainda incapazes de lidar razoavelmente com a linguagem humana.

A integração de dados clínicos textuais registrados em PEPs é um problema de pesquisa em aberto. Existem alternativas diversas e uma, em especial, que tem gerado bons resultados ao redor do mundo é a integração via ontologias (CEUSTERS; SMITH; FLANAGAN, 2003). Nesse contexto, o **objetivo** deste estudo é descrever aspectos necessários para o

entendimento do problema, além de fornecer alternativas para integrar dados clínicos em PEP com terminologias clínicas padronizadas no formato de uma metodologia.

A metodologia reflete a investigação realizada em um hospital brasileiro de grande porte, onde analisaram-se dados de milhares de registros médicos reais, confrontando-os com a classificação que receberiam em bem conhecidas terminologias clínicas (SOUZA; ALMEIDA, 2019; SOUZA, 2021). Para alcançar os objetivos, o restante do presente estudo se organiza da seguinte forma.

A Seção 2 visita as terminologias clínicas e soluções que envolvem ontologias, para então descrever o problema da sobreposição epistemológica em terminologias clínicas. A Seção 3 explica as dificuldades na busca pela interoperabilidade entre sistemas médicos, enquanto a Seção 4 introduz questões tecnológicas para a extração de dados clínicos de PEP. Seção 5 explica a metodologia de pesquisa desenvolvida no campo da Ciência da Informação para a extração de dados clínicos de PEP e integração com terminologias padronizadas.

2 | TERMINOLOGIAS CLÍNICAS

Existem diferentes tipos de sistemas de organização do conhecimento (SOCs) na área médica, que buscam o desenvolvimento e a melhoria dos recursos de informação e dados, por exemplo (SCHULZ; LÓPEZ-GARCÍA, 2015): os cabeçalhos de assunto, por ex. o *Medical Subject Headings* (MeSH); as classificações, por ex. a Classificação Internacional de Doenças (CID); as terminologias médicas, por ex. SNOMED-CT; as ontologias formais, por ex. aquelas desenvolvidas no consórcio *OBO Foundry*.

As terminologias clínicas são importantes por realizar mapeamento de termos e buscar a integração em sistemas de informação em saúde (SIs) (DALIANIS, 2018). Para que sejam de fato úteis aos SIs médicos, precisam ser multilíngues, estar alinhadas às práticas clínicas e possibilitar relatórios gerenciais necessários à administração (RECTOR, 1999). Três tipos de terminologias em saúde são bem conhecidos (SCHULZ *et al.* 2017):

a) Terminologias de interface: compreendem termos usados em textos clínicos, o conhecido “jargão médico”, que contem abreviaturas e acrônimos, além de exibir significados variáveis por usuário e por tempo.

b) Terminologias de referência: contém termos bem definidos, reconhecidos como “conceitos”, “classes”, ou “descritores”, definidos formalmente.

c) Terminologias de agregação: fazem uso de regras hierárquicas, classes e princípios de disjunção, dentre outros, sendo adotadas em análises estatísticas.

Uma terminologia de interface, como o PEP, se presta à assistência ao paciente e à comunicação entre equipe médica. Já uma terminologia de referência, como a SNOMED-CT, é uma extensa descrição de alternativas de diagnóstico; o *Medical Subject Headings* (MeSH) é utilizado para classificar artigos indexados, enquanto a UMLS mapeia terminologias. Uma

terminologia de agregação, como a CID, é um sistema de codificação de doenças. Dentre as terminologias clínicas, destacam-se os PEPs e as ontologias biomédicas.

2.1 Terminologias clínicas padronizadas: vocabulários e ontologias

Ontologias biomédicas, terminologias e vocabulários controlados são uma solução popularmente adotada em sistemas de informação para representar a informação médica (CEUSTERS; SMITH; FLANAGAN, 2003). Entretanto, existe demanda por terminologias de saúde como representações não ambíguas e não redundantes (ELKIN; TUTTLE, 2012).

Grande parte dos vocabulários aplicados ao registro de informações em saúde seguem arranjos na forma de classificações hierárquicas, agrupados muitas vezes em formatos diversos (Quadro 1).

| Vocabulário | Tipo | Propósito | Base | Comparação | Formalismo | Exemplo |
|---------------------------------------|-----------------------------|--|-----------------------------|----------------------------|---|-----------|
| Vocabulário composicional intencional | Terminologia | Recuperação de fatos | Conceitos | Cálculo, regras explícitas | Descrição lógica | OpenGalen |
| Vocabulário combinatório intencional | Nomenclatura | Recuperação de casos | Termos, conceitos | Multi hierarquias | Gramática semântica | SNOMED |
| Vocabulário disjuntivo extensional | Classificação e codificação | Avaliação estatística, partição de objetos reais | Objetos e grupos de objetos | Mono hierarquias | Árvores de decisão hierárquica e determinística | CID |
| Vocabulário associativo extensional | Tesaro | Recuperação de literatura | Tópicos, conceitos | Busca via palavra-chave | Abordagem tipo frames | MeSH |

Quadro 1 – Tipos de vocabulários, utilização e propósito

Fonte: adaptado de MASSAD *et al.* (2003, p. 50).

Tendo em vista o problema da organização do conhecimento na área biomédica, ontologias tem sido uma alternativa muito abordada. Originada da Filosofia, como o estudo da natureza do ser, o estudo em ontologias alcançou notoriedade na Ciência da Informação (CI) e Ciência da Computação (CC), como artefato de representação (FARINELLI; ALMEIDA, 2019). As ontologias são usadas em questões de interoperabilidade semântica e heterogeneidade terminológica. Esse uso decorre da capacidade em representar conhecimento livre de ambiguidades, a partir de compreensão consensual do significado dos termos (BITTNER; DONNELLY; WINTER, 2005; GANGEMI *et al.*, 2002).

Ontologias são utilizadas na padronização de vocabulários como alternativa às terminologias médicas, dada a expressividade de formalismos lógicos (CEUSTERS; SMITH; FLANAGAN, 2003; SMITH, 2008; SMITH; BROCHHAUSEN, 2010). Entretanto, da mesma forma, à medida que cresce o número de ontologias biomédicas, surge mais ambiguidades

e inconsistência. Assim, surgiu a iniciativa *The Open Biological and Biomedical Ontologies* (OBO Foundry¹) determinando as práticas para desenvolvimento de ontologias biomédicas interoperáveis, logicamente bem formadas e cientificamente precisas (SMITH *et al.*, 2007).

2.2 O problema da sobreposição em terminologias clínicas

Uma característica das terminologias clínicas, indesejada do ponto de vista da interoperabilidade, é a presença de aspectos epistêmicos. Rector (1999) a denomina “sobreposição epistemológica”, definida como a presença de informação adicional aos termos. Outrossim, observa-se, na descrição das classes, informação que reflete conhecimento ou ignorância sobre os mesmos.

As representações, tanto em nível ontológico quanto epistemológico, são importantes para o contexto clínico, mas o efeito da sobreposição não é desejado. As declarações epistemológicas são cruciais para o atendimento clínico, mas não constituem a axiomas usados em ontologias. Como as ontologias são instrumentos voltados para máquinas e, dessa forma, incapazes de apreender contexto ou lidar com aspectos subjetivos, a presença de aspectos epistêmicos é um obstáculo quando se busca por interoperabilidade automática entre sistemas.

De fato, uma ontologia formal pressupõe a inexistência de aspectos epistêmicos. Esse fenômeno impacta na CID, por exemplo, um vocabulário mandatário nas instituições de saúde brasileiras. A CID se baseia em um extenso esquema de código alfanumérico e em capítulos. Os capítulos contêm vários agrupamentos, que consistem em um conjunto de categorias. As categorias correspondem a um código com uma letra e dois dígitos, subdivididas em subcategorias. Nas subcategorias, a cada código de categoria, junta-se um ponto e um algarismo do .0 ao .9, que formam o código completo da doença.

Esse tipo de organização esquemática permite espaços para que, em futuras revisões, se mantenha a ordenação (Di NUBILA, 2007). São nas subcategorias CID que se encontra a variedade de informação que configura a sobreposição. Por isso, as subcategorias da CID têm sido denominadas declarações CID, uma vez que não podem ser confundidas com categorias. Para mais detalhes, vide Teixeira (2019), estudo que lista de 12 tipos mais comuns de sobreposição encontradas nas declarações CID.

3 | A INTEROPERABILIDADE NO DOMÍNIO DA SAÚDE

Alcançar interoperabilidade na saúde representa, em última instância, munir as pessoas com informação síncrona sobre sua saúde, permitindo agilidade e segurança na tomada de decisões. As pessoas assim, podem participar ativamente em seu próprio tratamento. Entretanto, o alcance dessa conquista implica superar obstáculos relacionados ao acesso, tratamento e recuperação eficiente da informação de saúde.

A investigação ocorre no campo de estudos sobre interoperabilidade, definida como

¹ Disponível em: <<http://www.obofoundry.org/>>

a capacidade de recursos computacionais distintos em acessar, trocar, integrar e usar dados, intra e extra organizacionalmente, provendo portabilidade e otimização da saúde, em escala mundial (HIMSS, 2021). No âmbito da saúde, os esforços empreendidos para prover interoperabilidade demandam o conhecimento e a distinção de um espectro de elementos heterogêneos, assim como das diversas formas de interoperabilidade, para que seja possível propor soluções eficientes e uso profícuo de informação. Pelo menos cinco tipos de heterogeneidade são observáveis (TEIXEIRA; EMYGDIO, 2020; SCHULZ *et al.*, 2017; SMITH, 2008).

- Recursos humanos: cooperam no macro-ambiente e no micro-ambiente;
- Recursos de comunicação humana, tais como, legislação, jargão e padrões terminológicos; recursos de comunicação tecnológica, tais como, linguagens de representação, modelos semânticos e padrões de interoperabilidade;
- Processos de negócios que estabelecem as diversas práticas de saúde;
- Recursos tecnológicos que sustentam tais práticas;
- Dados, gerados a partir da interação entre atores distintos.

Diversas organizações internacionais provêm padrões para uniformizar a comunicação no contexto médico, destacando-se dentre elas, a *International Organization for Standardization* (ISO), a *Health Level Seven International* (HL7), o *European Committee for Standardization* (CEN), a *International Health Terminology Standards Development Organisation* (IHSTDO), a *openEHR Foundation*, e o *IHE International Committee* (IHE). Ainda assim, são comuns casos de sobreposição, semântica e sintática, entre os diversos padrões.

A natureza diversificada destes padrões, a falta de conhecimento sobre a origem de cada um, e a demanda de uso concorrente, comprometem fortemente a busca de interoperabilidade (ACCESS CT, 2016). A distinção quanto à natureza destes padrões pode ser consultada em Teixeira e Emygdio (2020). Iniciativas usando ontologia que têm possibilitado sucesso em projetos de interoperabilidade semântica na área da saúde podem ser consultadas em Smith *et al.* (2007).

4 | QUESTÕES TECNOLÓGICAS: EXTRAÇÃO E INTEGRAÇÃO DE DADOS

A extração de dados em SIs é uma necessidade recorrente em diversas pesquisas científicas. Neste momento, apresenta-se o estudo de caso que contempla diversos passos metodológicos sobre extração de dados em uma terminologia de interface, ou seja, um sistema de informação clínica.

4.1 Sistemas de informação e sistemas de recuperação da informação

O termo “sistema de informação” também é às vezes usado em Ciência da

Informação (CI) para se referir a Sistemas de Recuperação de Informação (SRI) centrados em documentos, domínio de aplicação familiar às bibliotecas. Para Manning, Raghavan e Schütze (2008), a Recuperação de Informação (RI) é a procura por material, geralmente documentos; de natureza não estruturada, geralmente texto, que satisfaça às necessidades de informação em grandes coleções. Envolve o armazenamento, organização e pesquisa de coleções.

Em todos os casos, a tarefa é a mesma: extrair um conjunto de itens que o usuário deseja ter, separando-o daqueles que não deseja. Não se trata de tarefa simples, pois envolve não apenas os aspectos técnicos, mas também aspectos comportamentais e psicológicos do usuário para entender o que diferencia itens desejados dos não desejados. Na Computação, é focada em prover aos usuários um fácil acesso às informação de seu interesse (BAEZA-YATES;RIBEIRO-NETO, 2011).

4.2 Processamento de Linguagem Natural

Dentre as técnicas necessárias para a manipulação de dados clínicos, destaca-se o Processamento de Linguagem Natural (PLN), subárea da Inteligência Artificial (IA) que manipula a estrutura morfológica e sintática, bem como a semântica e pragmática (PEREIRA; MONTERO, 2012).

O objetivo é extrair informação, por exemplo, como o sentimento expresso em uma postagem em uma rede social, ideias e conceitos expressos em um artigo, ou até tratar os dados para extração de conhecimento. Na CI, a PLN busca soluções para problemas relacionados à recuperação da informação, sendo definido como “um conjunto de técnicas computacionais para a análise de textos em um ou mais níveis linguísticos, com o propósito de simular o processamento humano da língua” (FERNEDA 2003, p.82).

4.3 PLN aplicados aos textos clínicos

Este amplo contexto tecnológico aplica-se ao processamento de textos clínicos. A Extração de Informação (EI) visa encontrar informação específica em um documento ou em coleções de documentos (DALIANIS, 2018). Para elaborar estratégias de extração de dados utilizando abordagens tecnológicas, fez-se necessário analisar no texto os seguintes aspectos:

- **Segmentação:** uma das primeiras etapas do PLN, na qual se realiza a tarefa de separar as sentenças e as palavras (DALIANIS, 2018);
- **Tokenização:** a tarefa é detectar os limites do “*token*” e partes do discurso seja, a palavra que é analisada nas tarefas seguintes do processamento morfológico (IBM, 2018);
- **Processamento morfológico:** analisa as partes do discurso, como verbos, substantivos e adjetivos, dentre outros, no texto. (SANTOS *et al.*, 2015);
- **Siglas e abreviaturas:** verifica a presença de acrônimos, faz correção de erros

ortográficos, marca partes do discurso (DALIANIS, 2018);

- Stop words: termos frequentes no texto, mas de pouco significado (BLAKE, 2011, p.134).
- Extração de termos e relações semânticas: chamada de “análise de texto”, interpreta significados pelo uso de técnicas, como a detecção de negação, a extração de relações, dentre outras.

5 | METODOLOGIA DE PESQUISA

Apresenta-se aqui uma metodologia baseada em Souza (2021), de forma a exemplificar a experiência de extração de dados clínicos de PEPs para propósitos de integração. Os passos metodológicos são descritos no restante da presente seção.

5.1 A amostra

O pesquisador deve verificar na instituição a política para gestão da informação. Essa política, em geral, estabelece comitês de avaliação e medidas restritivas para impedir a exposição de dados indevidamente. Pode haver limitação da faixa temporal ou a partes do conteúdo, dentre outros. Mesmo em instituições médicas públicas que adotam política de dados abertos² podem ter restrições visto os riscos de exposição do paciente.

A definição do tamanho da amostra deve ser avaliada pelo pesquisador, por conveniência ou outro critério científico. Existem situações onde o tamanho da amostra depende de políticas institucionais e outros fatores relacionados à exposição de dados sensíveis. Entende-se por “dados sensíveis”, informação capaz de identificar pessoas e instituições, como: nome, idade, sexo, endereço, cpf, cnpj, etc.

5.2 Estratégia técnica para extração de dados

A extração de dados depende diretamente do tipo de tecnologia usada em seu armazenamento e organização. A informação, em geral, está registrada em repositórios digitais e é a partir daí que os dados serão consultados e extraídos para que o pesquisador tenha acesso.

Para a realização deste processo, o profissional inicia um mapeamento dos dados e suas relações, para que o conjunto desejado seja selecionado. Neste processo, usa-se via de regra um banco de dados estruturado, a linguagem SQL para consulta, e especifica-se a informação via filtros de seleção. Após o processo de definição da consulta, os dados selecionados devem ser exportados para acesso do pesquisador.

5.3 Escolha do formato de saída na exportação

Grandes conjuntos de dados estruturados são, na maioria das vezes, informação que depende de *software* específicos para manipulação. Porém, há situações onde os dados podem ser exportados para formatos livres, passíveis de leitura e manipulação por *software*

² <https://wiki.dados.gov.br/Politica-de-Dados-Abertos.ashx>

populares como editores de texto e planilhas eletrônicas.

Entende-se por “formatos abertos”, formatos passíveis de uso por *software* independentes de seus desenvolvedores e formatos proprietários. Pode-se citar alguns formatos abertos relevantes, como o CSV, o TSV, o TXT, o XML, o JSON, dentre outros. Há ainda formatos proprietários que estão direcionados a determinados *software* por questões de compatibilidade. Por exemplo, formatos de *backups* que são exportados por um banco de dados e que devem ser restaurados pelo mesmo.

5.4 Adequação do formato

Após exportados os dados, seja em formato aberto ou proprietário, o pesquisador pode desejar manuseá-los em *software* de sua preferência, a exemplo de planilhas eletrônicas. Neste sentido ocorre mais um processo de manipulação de dados, a saber, a importação para o *software* de escolha do usuário. Um exemplo de modelo de exportação é apresentado na Figura 1.

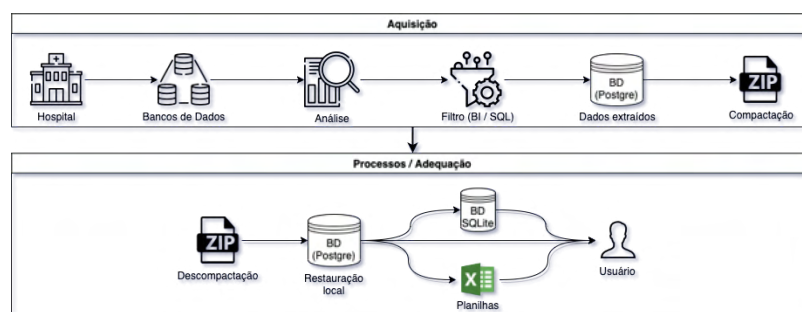


Figura 1 - Extração de dados do PEP.

Fonte: Adaptado de Souza (2021).

5.5 Comparação os dados extraídos e terminologias

Após a extração de dados clínicos do PEP, realiza-se a integração com a terminologia de referência e a terminologia de agregação. As ontologias biomédicas são recursos relevantes em tarefas de reconhecimento de entidades no texto para fins de extração de relações. Este tipo de processo é usado na técnica de mineração de texto, porque a ontologia define os tipos de entidades como as substâncias, qualidades e processos dos termos a relações entre eles (BODENREIDER, 2006).

Adotam-se as terminologias de referência (ontologias) do *OBO Foundry* para a integração com a terminologia de Interface, visto que estes artefatos seguem um padrão metodológico aceito pela comunidade científica. Também podem se utilizar a integração com terminologias de referência, vocabulários controlados da área de saúde, à exemplo do

MeSH, uma vez que tais artefatos apresentam sinônimos e definições textuais dos termos.

Na etapa da integração da terminologia de referência com a terminologia de agregação, utiliza-se a CID-10 por ser esta classificação utilizada na área de saúde no Brasil (Figura 2).

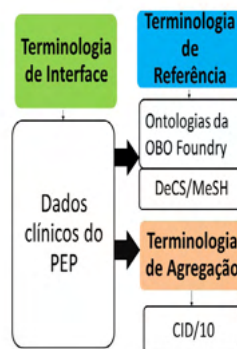


Figura 2 - Integração entre terminologias clínicas.

Fonte: Adaptado de Souza (2021).

6 | CONSIDERAÇÕES FINAIS

A heterogeneidade das terminologias médicas empregadas nos sistemas de saúde implica em desafios para a integração entre sistemas de saúde. A falta de padronização e sistematização dos dados quanto ao uso e definição de termos traz problemas para médicos e profissionais relacionados, além de dificultar a coleta e a extração de dados de PEPs.

A melhoria da interoperabilidade entre os SIs de saúde promove a troca de informação entre prontuários, possibilitando melhorias na qualidade dos cuidados de saúde prestados aos pacientes, reduzindo erros de diagnóstico e prescrição, além de minimizar a duplicidade de informação. As ontologias têm sido amplamente utilizadas na representação e organização de conhecimentos especializados em diversos campos da biomedicina.

A extração de dados clínicos é um processo que envolve várias etapas e artefatos tecnológicos que se complementam na realização deste objetivo. São muitos os pesquisadores que dependem desta informação para conduzir pesquisas. Existem também aspectos relacionados à legalidade, segurança e sigilo, além dos elementos tecnológicos fazem parte deste importante modelo de compartilhamento de informação. A integração terminológica é, de fato, um processo difícil que demanda diversas expertises para funcionar.

REFERÊNCIAS

ACCESS CT. **WP1 D1.4 Current and Future Use of SNOMED CT: Assessing SNOMED CT for Large Scale eHealth Deployments in the EU.** [s.l.] ASSESS CT, 2016. Disponível em: <<http://assess-ct.eu/>>. Acesso em: 13 mar. 2019.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval: the concepts and technology**

behind search. 2nd ed. New York: Addison Wesley, 2011.

BLAKE, C. Information retrieval. **ARIST**, v.45, n.1, sec. II, p.121-155, 2011.

BITTNER, T.; DONNELLY, M.; WINTER, S. **Ontology and semantic interoperability**. Large-scale 3D data integration: Challenges and Opportunities, p. 139-160, 2005.

BODENREIDER, O. Lexical, terminological and ontological resources for biological text mining. In: ANANIDOU, S. *et al.* **Text mining for biology and biomedicine**. Artech House: London, UK, 2006. p.43-66.

CEUSTERS, W.; SMITH, B.; FLANAGAN, J. **Ontology and medical terminology**: Why description logics are not enough. Towards an Electronic Patient Record (TEPR 2003), Boston, MA, 2003.

DALIANIS, H. Characteristics of Patient Records and Clinical Corpora. In: DALIANIS, H. **Clinical Text Mining: Secondary Use of Electronic Patient Records**. [s.n.], 2018b. cap. 4 Disponível em: <<http://link.springer.com/>>. Acesso em: 2 jan. 2019.

Di NUBILA, H.B.V. **Aplicações das classificações CID-10 e CIF nas definições de deficiência e incapacidade**. Tese (Doutorado). Faculdade de Saúde Pública, Universidade de São Paulo, 2007. Disponível em: <<http://www.teses.usp.br/>>. Acesso em: 21 ago. 2017.

ELKIN, P. L.; TUTTLE, M. S. History of terminology and terminological logics. In: ELKIN, P. L. (ed.). **Terminology and terminological systems**. London: Springer-Verlag, 2012. cap. 2, p.5-10.

FARINELLI, F.; ALMEIDA, M. B. Ontologias biomédicas: teoria e prática. In: ZIVIANI, A., FERNANDES, N.C.; SAADE, D.C.M. In: **18º Simpósio Brasileiro de Computação Aplicada à Saúde**, 2019, Niterói. Sociedade Brasileira de Computação, p.93-140.

FERNEDA, E. **Recuperação de Informação**: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 137f. Tese (Doutorado em Ciência da Informação e Documentação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.

GANGEMI, A.; *et al.* A formal ontological framework for semantic interoperability in the fishery domain. In: EUZENAT, J.GOMEZ-PEREZ, A.GUARINO, N., *et al.*, **European Conference on Artificial Intelligence (ECAI'02) - Workshop on Ontologies and Semantic Interoperability**, 2002, Lyon, France. CEUR Workshop Proceedings, 64, July 22. p.16-30.

HEALTHCARE INFORMATION AND MANAGEMENT SYSTEMS SOCIETY (HIMSS). **Interoperability in HealthCare**. Disponível em: <https://www.himss.org/resources/interoperability-healthcare#Part1>. Acesso: 18, mar. 2020.

IBM. Knowledge Center. **Tokenização**. Disponível em; <https://www.ibm.com/>. Acesso em 08 de nov 2018.

MANNING, C.D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. New York: Cambridge University Press, 2008.

MASSAD, E.; (ed.) *et al.* **A. O prontuário eletrônico do paciente na assistência, informação e conhecimento médico**. São Paulo: H. de F. Marin, 2003. 213p.

PEREIRA, T.A.; MONTERO, E.F.S. Terminologia DeCS e as novas regras ortográficas da língua portuguesa: orientações para uma atualização. **Acta Cir Bras**, v. 27, n. 7, p. 509–514, jul. 2012.

RECTOR, A. L. Clinical terminology: why is it so hard? **Methods Inf Med**, v. 38, n. 4, p. 239-252, 1999.

REIS, Z.S.N.; CORREIA, R.J.C.; PEREIRA, A.D.C. Sistemas eletrônicos de informação na assistência e pesquisa em saúde da mulher: para quando um maior envolvimento dos profissionais de saúde. **Rev Bras Ginecol Obstet**, v. 33, n. 3, p. 107-10, 2011. Disponível em: < <http://www.scielo.br/pdf/rbgo/v33n3/a01v33n3.pdf> >.

SANTOS, R.E.S.; *et al.* Técnicas de processamento de linguagem natural aplicadas ao processo de mineração de textos: resultados preliminares de um mapeamento sistemático. **Revista de Sistemas e Computação-RSC**, v.4, n.2,2015.

SCHULZ, S. *et al.* Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Stud Health Technol Inform**, Amsterdam, v.245, p.940-944, 2017.

SCHULZ, S.; LÓPEZ-GARCÍA, P. Big Data, medizinische Sprache und biomedizinische Ordnungssysteme. **Bundesgesundheitsbl**, v. 58, n. 8, p. 844–852, 2015. Disponível em: < <https://link.springer.com> >. Acesso em ago 2021.

SMITH, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nat Biotechnol**, v. 25, n. 11, p. 1251–1255, nov. 2007.

SMITH, B. New desiderata for biomedical terminologies. In: MUNN, K.; SMITH, B. **Applied Ontology: an Introduction**. Frankfurt: Ontos Verlag, 2008.p.83-107, cap.4.

SMITH, B.; BROCHHAUSEN, M. Putting biomedical ontologies to work. **Methods Inf Med**, v. 49, n. 2, p. 135, 2010. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116518/pdf/nihms296603.pdf>>. Acesso em ago 2021.

SMITH, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nat Biotechnol**, v. 25, n. 11, p. 1251-1255, 2007.

SOUZA, A.D.; ALMEIDA, M.B. Integração de dados clínicos textuais de Prontuários Eletrônicos do Paciente com terminologias médicas padronizadas. In: ALMEIDA, J.A.P. *et al.* **Proceedings of the XII Seminar on Ontology Research in Brazil**. Porto Alegre, Brazil, September 2nd-5th, 2019.

SOUZA, A.D. **O discurso na prática clínica e as terminologias de padronização**: investigando a conexão. 2021. 286f. Tese (Doutorado em Gestão e Organização do Conhecimento) Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2021.

TEIXEIRA, L.M.D.; EMYGDIO, J.L. Realismo ontológico aplicado à interoperabilidade semântica na saúde In: ALMEIDA, M.B. (ed). **Representação do conhecimento, ontologias e linguagem**: pesquisa aplicada em Ciência da Informação. Curitiba: CRV, 2020. p. 47-82.v.I.

TEIXEIRA, L.M.D. **Princípios ontológicos aplicados à classificação Internacional de doenças**: alternativas para a busca por interoperabilidade semântica entre sistemas de prontuários eletrônicos de pacientes. Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação. Belo Horizonte, 2019.

