



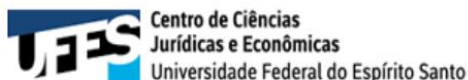
5º WIDaT

ANAIS do Workshop de Informação, Dados e Tecnologia

05 e 06 de dezembro/2022

UFES - Campus Goiabeiras
Auditório da Biblioteca Central

Realização:



Apoio:



ISBN: 978-65-00-61990-4

WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA
Universidade Federal do Espírito Santo – UFES
05 e 06 de dezembro de 2022, Vitória – ES

ANAIS WIDAT 2022
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Organizadores
Henrique Monteiro Cristovão (PPGCI/UFES)
Daniela Lucas da Silva Lemos (PPGCI/UFES)

Vitória-ES
2023

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

Workshop de informação, dados e tecnologia (5. : 2022
dez. : Vitória).

W924 Workshop de informação, dados e tecnologia (WIDAT
2022): anais do V WIDaT [recurso eletrônico] / Henrique Monteiro
Cristovão, Daniela Lucas da Silva Lemos, organizadores. - Dados
eletrônicos. - Vitória : PPGCI/UFES, 2022.
247 p. ; il.

Workshop realizado no período de 5 e 6 de dez. de 2022.
Inclui bibliografia.
ISBN: 978-65-00-61990-4
Modo de acesso: <https://widat2022.ufes.br/anais>

1. Ciência da informação - Workshop. 2. Tecnologia.
I. Henrique Monteiro Cristovão. II. Daniela Lucas da Silva Lemos.
III. Título.

CDU: 02

Organização WIDaT 2022

Coordenação Geral:

Daniela Lucas da Silva Lemos (PPGCI/UFES)

Comissão Organizadora:

Dalton Lopes Martins (PPGCI/UNB)

Gleice Pereira (PPGCI/UFES)

Henrique Monteiro Cristovão (PPGCI/UFES)

Marta Leandro da Mata (PPGCI/UFES)

Morgana Andrade (Biblioteca Central/UFES)

Paula Regina Ventura Amorim Gonzalez (Depto. Biblioteconomia/UFES)

Coordenação da Comissão Científica:

Henrique Monteiro Cristovão (PPGCI/UFES)

Comissão Científica:

Alisson Marques da Silva (CEFET-MG)

Ana Carolina Simionato (PPGCI/UFSCar)

Dalton Martins (PPGCI/UNB)

Daniela Lucas da Silva Lemos (PPGCI/UFES)

Diana Vilas Boas Souto Aleixo (DARQ/UFES)

Douglas Dyllon Jerônimo de Macedo (CIN/UFSC)

Eduardo Ribeiro Felipe (UNIFEI)

Fabício Mendonça (DCC/ UFJF)

Fernanda Farinelli (PPGCI/UNB)

Jeanne Louize Emygdio (PUC-MG)

Jose Eduardo Santarem Segundo (PPGCI/UNESP/ USP)

Luciana Itida Ferrari (DARQ/UFES)

Marta Leandro da Mata (PPGCI/UFES)

Michel Pires (CEFET-MG)

Moisés Lima Dutra (CIN/UFSC)

Morgana Andrade (Biblioteca Central /UFES)

Patrícia Dias (UEMG)

Paula Regina Ventura Amorim Gonzalez (DBIB/UFES)

Renato Rocha Souza (CPDOC/FGV)

Sandro Rautenberg (PGCIN-UFSC)

Tânia Barbosa Salles Gava (DARQ/UFES)

Thiago Magela Rodrigues Dias (CEFET-MG)

Tiago Alves de Oliveira (CEFET-MG)

Washington Segundo (IBICT)

Comissão Técnica:

Bruna Stefane de Freitas (UFES)

Calíope Victor Spíndola de Miranda Dias (UnB)

Dirceu Flávio Macedo (Grupo TOIC/UFES)

Kris Ellen das Neves Teixeira (UFES)

Luísa Vernersbach Varejão (UFES)

Rafael Denerson Ramos de Sousa (UFES)

Silvana Pires Rocha Nogueira (UFES)

Sumário

A APLICABILIDADE DA FOLKSONOMIA NOS ESTUDOS ALTMÉTRICOS	9
Skrol Salustiano Fabio Castro Gouveia	
A INDÚSTRIA 4.0 SOB O PRISMA DA CIÊNCIA DE DADOS: UMA PROPOSTA DE MODELO DE MATURIDADE	15
Jacqueline Zonichenn Reis	
A RELAÇÃO DA ORGANIZAÇÃO DO CONHECIMENTO COM AS HUMANIDADES DIGITAIS A PARTIR DAS DEZ PREMISSAS DE BARITE	21
Ana Cristina de Albuquerque Marcos Antonio de Moraes Ania Rosa Hernandez Quintana	
A WEB SEMÂNTICA APLICADA NA RECUPERAÇÃO DE INFORMAÇÃO: UM ESTUDO DE CASO NO CONTEXTO ESTATÍSTICO DE USO DE LIVROS DIGITAIS POR ALUNOS DE GRADUAÇÃO	29
Stella Schwanz Dias de Assis Alessandra Monteiro Pattuzzo Caetano Henrique Monteiro Cristovão	
ANÁLISE DA REGIONALIDADE DO CONJUNTO DE ARTIGOS PUBLICADOS EM EVENTOS CIENTÍFICOS	36
Fernanda Silva Coimbra Thiago Magela Rodrigues Dias Ronaldo Ferreira de Araújo	
BRASILIANA MUSEUS: TESTE FUNCIONAL DO AGREGADOR DE DADOS MUSEAIS DO INSTITUTO BRASILEIRO DE MUSEUS	42
Joyce Siqueira Dalton Lopes Martins Vinícius Nunes Medeiros	
CIÊNCIA DE DADOS: UMA REVISÃO DE LITERATURA EM PERIÓDICOS DA CIÊNCIA DA INFORMAÇÃO	48
Aurea Celeste Pires de Souza Clarice Luzia Casoni Merabe Carvalho Ferreira da Gama José Eduardo Santarem Segundo	
CLASSIFICAÇÃO AUTOMÁTICA DE ARTIGOS PUBLICADOS NO ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - 2021	57
Liliane Cristina Soares Sousa José Eduardo Santarem Segundo Fábio Parra Furlanete	

CONTRATOS INTELIGENTES NO DESENVOLVIMENTO DE COLEÇÕES: UMA ABORDAGEM ORIENTADA À BLOCKCHAIN	63
Rafael Rocha	
Gercina Ângela de Lima	
CRIAÇÃO E CAPTURA DE VALOR BASEADAS EM BIG DATA PARA A INOVAÇÃO EM PRODUTOS E SERVIÇOS: ANÁLISE DA PRODUÇÃO CIENTÍFICA	69
Priscila Machado Borges Sena	
Nathalia Berger Werlang	
Ana Clara Cândido	
DESCOBERTA DE CONHECIMENTO APLICADA À BASE DE DADOS ABERTOS DA ANVISA SOBRE PREÇOS DE MEDICAMENTOS POR MEIO DE ANÁLISE DE REDES DE INFORMAÇÃO	75
Lucas Vale	
Henrique Monteiro Cristovão	
DESCOBERTA DE RELAÇÕES ENTRE ESTADOS BRASILEIROS A PARTIR DE DADOS FINANCEIROS DE OPERAÇÕES DE CRÉDITO DISPONÍVEIS EM DADOS ABERTOS DO BANCO CENTRAL	83
Juliana Rodrigues de Lima Meirelles	
Henrique Monteiro Cristovão	
Daniela Lucas da Silva Lemos	
DETECÇÃO DE VAZAMENTOS DE FLUIDOS DE FREIOS A AR EM VAGÕES DO TIPO GÔNDOLA ATRAVÉS DO SINAL ACÚSTICO: UM MODELO DE CLASSIFICAÇÃO DE FALHAS	91
Jordana Lucia Reis	
Flávio Miguel Varejão	
DIAGNÓSTICO DE FALHAS: UMA REVISÃO E ANÁLISE DE DADOS DE VIBRAÇÃO E SUAS APLICAÇÕES	97
Igor Varejão	
Alexandre Rodrigues Loureiros	
Thiago Oliveira-Santos	
Flávio Varejão	
DIREITOS AUTORAIS RELACIONADOS À MEMÓRIA INSTITUCIONAL E ARTÍSTICA DO TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E DOS TERRITÓRIOS....	103
Rosilene Paiva Marinho de Sousa	
Maison Roberto M. Gonçalves	
Diego José Macedo	
Milton Shintaku	
HL7 FHIR BASEADO EM W3C PROV PARA ALCANÇAR A PROVENIÊNCIA DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE	108
Márcio José Sembay	
Douglas Dyllon Jeronimo de Macedo	
Alexandre Augusto Gimenes Marquez Filho	

IMAGO: UMA PROPOSTA PARA O BANCO DE IMAGENS DO INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA	115
Diego José Macedo	
Ítalo Barbosa Brasileiro	
Milton Shintaku	
IMPACTO DA ADEQUAÇÃO À LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS NA METRIFICAÇÃO DA QUALIDADE DE DADOS	121
Leandro Furlam Turi	
Giovanni Comarela	
LIBERDADE DE EXPRESSÃO: A NARRATIVA NO TWITTER EM UM CONTEXTO DE ANÁLISE DE REDES SOCIAIS	127
Stella Schwanz Dias de Assis	
Meri Nadia Marques Gerlin	
METADADOS PARA COLEÇÕES E ACERVOS ARTÍSTICOS UNIVERSITÁRIOS	134
Aline Cristina Gomes Ramos	
Daniela Lucas da Silva Lemos	
MÉTODOS ÁGEIS NA CIÊNCIA DA INFORMAÇÃO: ENSINO DO SCRUM	141
Patrícia Nascimento Silva	
MODELO DE PRESERVAÇÃO HIPATIA: METODOLOGIA DE ESTUDO DE METADADOS PARA EXTRAÇÃO	147
Ívina Flores Melo	
Tatiana Canelhas	
Tiago Emmanuel Nunes Braga	
ORGANIZAÇÃO DE OBJETOS DE APRENDIZAGEM COM BASE NO PADRÃO LOM-IEEE UTILIZANDO DADOS LIGADOS INTEROPERÁVEIS NA WEB SEMÂNTICA	153
Viviane Bessa Lopes Alvarenga	
Henrique Monteiro Cristovão	
PATENTES COMO FONTE DE DADOS PARA ANÁLISE SOBRE A PRODUÇÃO TÉCNICA	160
Raulivan Rodrigo da Silva	
Thiago Magela Rodrigues Dias	
PLATAFORMA PARA AGREGAÇÃO E ANÁLISES DE DADOS TÉCNICOS-CIENTÍFICOS	166
Thiago Magela Rodrigues Dias	
Washington Luís R. de Carvalho Segundo	
Tales Henrique José Moreira	
Vivian dos Santos Silva	
Adilson Luiz Pinto	

PREVISÃO DE CRIMES EM VALÊNCIA: UMA ABORDAGEM MULTI-RÓTULO	171
Luís Eduardo Freire da Câmara	
Alexandre Rodrigues Loureiros	
Jorge Mateu Mahiques	
Flávio Miguel Varejão	
PROCESSAMENTO DE LINGUAGEM NATURAL E MACHINE LEARNING COMO APARATO PARA A CATEGORIZAÇÃO DE ARTIGOS CIENTÍFICOS	177
Ananda Fernanda de Jesus	
Maria Lígia Triques	
José Eduardo Santarem Segundo	
Ana Cristina de Albuquerque	
PROCESSO SISTEMÁTICO FUNDAMENTADO EM MODELAGEM ONTOLÓGICA APLICADO À ESTRATIFICAÇÃO DE RISCO EM SAÚDE MENTAL PARA ANÁLISE QUALI-QUANTI	183
Evaldo de Oliveira da Silva	
Yuri Bento Marques	
Marcello Peixoto Bax	
PRODUÇÃO CIENTÍFICA EM CIÊNCIA DA INFORMAÇÃO: UTILIZANDO OS DADOS ABERTOS CAPES.....	190
Patrícia Ofélia Pereira de Almeida	
Patrick Stacy Meyer	
QUALIFICAÇÃO DE REPOSITÓRIOS DE DADOS E DE PUBLICAÇÕES: UMA PROPOSTA DE CRITÉRIOS ALINHADA À CIÊNCIA ABERTA	195
Priscila Machado Borges Sena	
Tatyane Guedes Martins da Silva	
Juliana Araujo Gomes de Sousa	
Washington Luís Ribeiro de Carvalho Segundo	
Bianca Amaro de Melo	
REPOSITÓRIOS DE DADOS DE PESQUISA: ANÁLISE À LUZ DOS PRINCÍPIOS FAIR	200
Letícia Guarany Bonetti	
Ana Carolina Simionato Arakaki	
REPRESENTAÇÃO DA INFORMAÇÃO EM MUSEUS: UMA REVISÃO SISTEMÁTICA EM BASES DE DADOS BRASILEIRAS	207
Bruna Stefane de Freitas	
Kris Ellen das Neves Teixeira	
Luísa Vernersbach Varejão	
Silvana Pires Rocha Nogueira	
Daniela Lucas da Silva Lemos	
Dalton Lopes Martins	

REPRESENTAÇÃO DE DADOS DE PESQUISA COM O PADRÃO DUBLIN CORE: UMA PROPOSTA DE MODELAGEM DE METADADOS PARA PROJETOS COVID-19 NO ÂMBITO DA UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	214
Beatriz Mantovani Avancini de Jesus Daniela Lucas da Silva Lemos Nádia Elôina Barcelos Fraga	
RISCOS E OPORTUNIDADES PARA OS USUÁRIOS DAS FINANÇAS DESCENTRALIZADAS	222
Fábio Cossenzo Marcello Peixoto Bax	
TERMOS EM MOVIMENTO: DINÂMICA DA REDE TEMÁTICA NO PROJETO BIBLIOTECA COMUM	229
Benjamin Luiz Franklin	
TRATAMENTO DA INFORMAÇÃO EM ACERVOS CULTURAIS: AVALIAÇÃO DO USO DE VOCABULÁRIOS CONTROLADOS EM COLEÇÕES MUSEOLÓGICAS SOB GESTÃO DO INSTITUTO BRASILEIRO DE MUSEUS	235
Abeil Coelho Júnior Daniela Lucas da Silva Lemos	
VISUALIZAÇÃO DE DADOS ABERTOS NO CONTEXTO DA PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO: ANÁLISE BIBLIOMÉTRICA DOS ESTUDOS DEFENDIDOS	242
Francis Bento Marques Yuri Bento Marques Benildes Coura Moreira dos Santos Maculan Renato Rocha Souza	

PROCESSO SISTEMÁTICO FUNDAMENTADO EM MODELAGEM ONTOLÓGICA APLICADO À ESTRATIFICAÇÃO DE RISCO EM SAÚDE MENTAL PARA ANÁLISE QUALI-QUANTI

SYSTEMATIC PROCESS BASED ON ONTOLOGICAL MODELING APPLIED TO RISK STRATIFICATION IN MENTAL HEALTH FOR QUALI-QUANTITY ANALYSIS

Evaldo de Oliveira da Silva, Yuri Bento Marques, Marcello Peixoto Bax¹

(1) Programa de Pós-Graduação em Ciência da Informação – UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, evaldosilva@ufmg.br, yuri.marques@ifnmg.edu.br, bax@eci.ufmg.br

Resumo

Transtornos mentais estão entre as causas de inaptidão no mundo. A ansiedade é um transtorno causado por trauma, estresse devido a uma doença ou a partir de outras desordens mentais. Estudos de saúde mental frequentemente utilizam estimativas e dados sociodemográficos e sobre psicopatologias. Utilizam-se instrumentos de avaliação que permitem tabular valores referentes aos sinais e sintomas e acompanhá-los na evolução do tratamento, como a estratificação de risco. Dados estruturados e não estruturados são gerados a partir desses instrumentos, o que dificulta as análises quantitativas e qualitativas, caso os sinais e sintomas sejam reavaliados ou se a eficácia e eficiência dos tratamentos serviram para mitigar o risco daquele paciente. Desta forma, este artigo apresenta um processo sistemático fundamentado na modelagem ontológica para análise quantitativa e qualitativa de dados da estratificação de risco em ansiedade, em conjunto com dados extraídos dos tratamentos psicoterápicos. O processo aborda o uso de ontologias e a técnica SDD (*Semantic Data Dictionary*) para compreensão e preparação dos dados. São gerados grafos de conhecimento que contêm os dados anotados semanticamente. Esses grafos podem ser úteis para encontrar relações entre doenças, sintomas e tratamentos além de servir como ferramenta para formação de diagnóstico e auxílio na definição dos níveis de cuidados com o paciente.

Palavras-chave: Saúde Mental, Ontologia, Psicoterapia, Análise Quali-Quanti, Grafos de Conhecimento.

Abstract

Mental disorders are among the leading causes of disability in the world. Anxiety is a disorder caused by trauma, stress due to an illness, or from other mental disorders. Mental health studies often use estimates and socio-demographic and psychopathology data. Assessment tools are used to chart values for signs and symptoms and track them during treatment, such as risk stratification. Structured and non-structured data are generated from these instruments, which makes quantitative and qualitative analyses difficult, should signs and symptoms be reevaluated or if the effectiveness and efficiency of treatments have served to mitigate that patient's risk. Thus, this article presents a systematic process based on ontological modeling for the quantitative and qualitative analysis of risk stratification data in anxiety, together with data extracted from psychotherapeutic treatments. The process addresses the use of ontologies and the Semantic Data Dictionary (SDD) technique for data understanding and preparation. Knowledge graphs are generated and hold the semantically annotated data. These graphs can be useful for finding relationships between diseases, symptoms and treatments as well as serving as a tool for forming diagnoses and helping to define levels of patient care.

Keywords: Mental Health, Ontology, Psychotherapy, Quali-Quantity Analysis, Knowledge Graphs.

1. Introdução

Transtornos mentais acometem milhões de pessoas no mundo que sofrem de depressão, ansiedade e cometem suicídios (WHITEFORD, FERRARI e DEGENHARDT, 2016). A ansiedade é um transtorno mental causado por trauma, estresse devido a uma doença, ou ainda, a partir de outros transtornos mentais. Dados permitem análise da condição mental dos pacientes,

por exemplo, instrumentos de avaliação contendo estruturas que permitem tabular escalas de valores referentes aos sinais e sintomas observados e os textos sobre a evolução dos tratamentos (MELLO, 2015; PAULA, 2019). Instrumentos de avaliação usados em estratificação de risco, revisam metas e orientam diferentes intervenções, conforme as necessidades dos pacientes (BRASIL, 2014).

A estratificação de riscos é um processo que utiliza instrumentos de avaliação no formato de questionário, que é aplicado para atribuir uma faixa de risco ao paciente com base em uma pontuação final. Equipes multiprofissionais utilizam o risco para estabelecer metas dos tratamentos dos planos de intervenção com o objetivo de mitigar o risco do paciente (MELLO, 2015; (PAULA, 2019; PARANÁ, 2021). Esses conjuntos de dados (estratificação e tratamentos), podem estar em formatos distintos (estruturados ou formatos textuais) dificultando as análises da eficácia dos tratamentos, se os riscos foram mitigados ou quais fatores que influenciam no tratamento. Além disso, estudos científicos podem ser desenvolvidos com base nos conjuntos de dados mencionados.

A *data science* refere-se à aplicação de métodos de obtenção de resultados científicos por meio da utilização de computação, e grande volume de dados (BERTHOLD *et al.*, 2010; SCHRÖER, KRUSE e GÓMEZ, 2021). Possui um processo suportado pelas fases iniciais de compreensão e preparação de dados.

A compreensão de dados visa entender se existem dados suficientes para responder a diferentes questões de pesquisa (BERTHOLD *et al.*, 2010). Alguns trabalhos utilizam ontologias como mecanismos para compreensão de dados atribuindo-lhes significados (semântica) (BRISSON e COLLARD, 2008; SVÁTEK, RAUCH e RALBOVSKÝ, 2005). Ademais, pesquisas apresentam o uso de ontologias para a organização do conhecimento na saúde mental (BRENAS, SHIN e SHABAN-NEJAD, 2019; CEUSTERS e SMITH, 2010; HASTINGS, 2012; YAMADA, 2020).

A preparação de dados abrange a seleção e mapeamento dos conjuntos de dados para suportar as técnicas de modelagem evitando análises enviesadas (PATON, 2019).

Dicionários de dados e metadados facilitam o compartilhamento, classificação e o mapeamento de dados na área da saúde e oferecem apoio para a preparação de dados (AHIMA, 2022). O ONC¹ tem desenvolvido

projetos para promover padrões de metadados e dados de saúde, que suportam os princípios localizáveis, acessíveis, interoperáveis e reutilizáveis (ou FAIR²) (HEALTHIT, 2020; ZAYAS-CABÁN e WALD, 2020).

O uso de SDD (*Semantic Data Dictionary*) tem sido aplicado em consonância com os princípios FAIR para diferentes estudos científicos. Rashid *et. al* (2017) utilizam padrões de metadados para configurar a anotação semântica usando SDD. Os autores recomendam a utilização da ontologia de topo SIO (*Semanticscience Integrated Ontology*) que fornece propriedades para descrever os relacionamentos entre objetos e atributos como modelo de representação do conhecimento na área biomédica.

Diante dos diferentes tipos de dados gerados a partir dos instrumentos de avaliação e tratamento dos casos clínicos, a seguinte questão é apresentada: "Como compreender dados estruturados e não estruturados a fim de prepará-los para análise quali-quantitativa?". Desta forma, este artigo apresenta um processo sistemático fundamentado em modelagem ontológica aplicado à estratificação de risco em saúde mental e tratamentos psicoterápicos para análise quali-quantitativa.

O restante do artigo está organizado da forma que segue. A Seção 2 descreve a fundamentação teórica. A Seção 3 estabelece quais procedimentos metodológicos são utilizados para o processo proposto. A Seção 4 apresenta os resultados alcançados. A Seção 5 descreve as considerações finais.

2. Fundamentação teórica

2.1. Dicionário Semântico de Dados

Rashid *et. al* (2017) usam padrões de metadados para configurar a anotação semântica por um SDD. A anotação semântica proposta por Rashid *et. al* (2017) utiliza os seguintes documentos:

- *InfoSheet*: referências para descrição dos SDDs;

¹ Office of the National Coordinator for Health Information Technology

² Findability, Accessibility, Interoperability, e Reuse

- *Dictionary Mapping*: anotação semântica das colunas das coleções de dados;
- *CodeBook*: códigos correspondentes a conceitos de ontologia;
- *Code Mapping*: mapeamento de termos dos datasets que correspondem a conceitos existentes na ontologia;
- *TimeLine*: anotação de intervalos temporais;
- *Properties Table*: para fins de customizar a descrição por outras ontologias de topo.

A ferramenta *sdd2rdf*³ interpreta o SDD e processa os dados, formando um grafo RDF. Para acessar os dados anotados, o *sdd2rdf* cria consultas no formato SPARQL. São geradas também regras SWRL⁴ que auxiliam em novas inferências.

A técnica SDD é utilizada neste artigo visando explicitar os conceitos sobre os dados para gerar grafos de conhecimento dos dados anotados semanticamente. Os grafos serão utilizados para encontrar relações entre doenças, sintomas e tratamentos, servindo para auxiliar na definição dos níveis de cuidados com o paciente.

2.2. Análise qualitativa apoiada por software

Análises qualitativas podem ser apoiadas por softwares (CAQDAS – *Computer Assisted Qualitative Data Analysis*). A tarefa de *coding* permite definir ou categorizar os dados (textuais) que são analisados para diferentes objetivos de pesquisas. Essa tarefa utiliza de procedimentos que revelam temas embutidos nos dados (WILLIAMS e MOSER, 2019). CAQDASs organizam os *codings* em estruturas baseadas em desenhos qualitativos, que servem de base para codificar (anotar) os textos dos tratamentos psicoterápicos.

2.3. O framework FLAIR

Akbik *et al.* (2019) apresentam o FLAIR⁵, um framework para processamento de linguagem natural (NLP, *Natural Language*

Processing), utilizado para facilitar a classificação de texto. O FLAIR permite que o pesquisador aplique modelos de NLP para reconhecer NER (*Named Entity Recognition*), PoS (*Part-of-Speech Tagging*), com suporte especial para dados biomédicos, desambiguação e classificação de termos. NLP é importante no contexto deste trabalho, para extração de termos que evidenciem as ações terapêuticas, técnicas aplicadas para mitigação de riscos e termos biomédicos existentes nos tratamentos psicoterápicos.

2.4. Trabalhos Relacionados

Em Han e Stoffel (2011) os dados textuais são anotados com ontologias e as informações são recuperadas de diferentes maneiras a fim de permitir que pesquisadores possam aprender e inferir novos conhecimentos.

Kapiszewski e Karcher (2021) destacam que estudos científicos devem envolver a transparência da pesquisa qualitativa, anotação, exportação de códigos por meio de CAQDAS, bem como o compartilhamento de dados. Além disso, devem conter informações referentes à descrição da análise dos dados, qual método foi utilizado e se os *codings* foram feitos de forma indutiva ou dedutiva.

Hocker *et al.* (2021) discutem a troca de conhecimento dos dados analisados qualitativamente. De acordo com os autores, o *coding* deve fornecer uma documentação formal para suportar o compartilhamento dos dados. A ontologia QualiCO, desenvolvida pelos autores supracitados, visa preencher esta lacuna com o uso de metadados. Os metadados permitem que pesquisadores anotem informações sobre os *codings* a fim de enriquecer o conhecimento sobre os dados por eles categorizados (HOCKER *et al.*, 2021).

3. Processo sistemático fundamentado em modelagem ontológica para análise quali-quantitativa

O processo apresentado neste trabalho é fundamentado no uso de ontologias para compreender e preparar dados para análise quali-quantitativa. O processo é definido em 5 etapas a seguir:

³ <https://github.com/tetherless-world/SemanticDataDictionary>

⁴ <https://www.w3.org/Submission/SWRL/>

⁵ <https://github.com/zalandoresearch/flair>

1. Modelagem ontológica. Criação/ajuste de ontologia de domínio (fundamentada em ontologias de topo) para formalização dos conceitos tratados no problema de pesquisa, incluindo a reutilização de ontologias consolidadas no domínio do problema.

2. *Coding*. O processo utiliza CAQDAS para análise qualitativa de dados não estruturados. O *coding* é construído a partir dos conceitos existentes na ontologia. Profissionais especialistas do domínio ou pesquisadores devem participar da modelagem ontológica com o objetivo de consensuar o conhecimento para análise qualitativa.

3. Execução de NLP. Ao terminar a análise qualitativa, os *codings* são extraídos para serem armazenados em *datasets* com os respectivos textos codificados. Então, NLP é executado em cada *coding* extraído, a fim de reconhecer elementos textuais, tais como, sujeito, ação, complemento da ação, doença e informações adicionais.

4. Unificação da estrutura de dados. Os dados reconhecidos por NLP são organizados em conjunto com dados quantitativos em um mesmo formato tabular, a fim de facilitar a anotação semântica.

5. Anotação Semântica. A anotação baseia-se em Rashid et. al (2017). Permite gerar o grafo RDF (*script sdd2rdf*) persistido no banco de dados *triplestore*. A ontologia formaliza o vocabulário. Após definir o *dataset* para anotar, os seguintes artefatos são gerados:

- *Dictionary Mapping* (DM). Cada linha do *DM* mapeia uma coluna do dataset, formalizando-a conceitualmente e também suas relações e proveniências.
- *CodeBook*. Permite a criação dos seguintes campos: Coluna (entidade a ser anotada), Código, Descrição e a Classe da Ontologia.
- Grafo RDF. Processamento dos artefatos "SDD + Dados" pelo *script sdd2rdf*, gerando o RDF e armazenando-o em um banco de dados *triplestore* para consulta posterior.

Os dados dos objetos mapeados pelo SDD são as colunas do próprio *dataset*. Porém, Rashid et al. (2017) afirmam que os objetos descritos no *dataset* podem encontrar-se ali explícita ou implicitamente. Ou seja, no

mesmo *dataset* podem ser gerados novos atributos provenientes de outros objetos implicitamente representados. Estes objetos serão explicitados no SDD e formalizados no grafo final gerado (pelo script *sdd2rdf*), favorecendo a sua integração nos níveis conceituais (ou intencionais) mais abstratos do projeto.

4. Resultados

Descreve-se um processo sistemático fundamentado em modelagem ontológica para análise quali-quantitativa aplicada à estratificação de risco em saúde mental e tratamentos psicoterápicos (representado na Figura 1). Os dados da estratificação se encontram em formato tabular. As etapas do processo seguem da forma que são executadas:

3. Modelagem ontológica. A ontologia MHMO⁶ (*Mental Health Management Ontology*) auxilia na gestão das redes de saúde mental e modela diferentes transtornos mentais.

4. *Coding*. Utiliza-se como CAQDAS a ferramenta Atlas-Ti⁷ para análise qualitativa dos dados dos tratamentos psicoterápicos. O *coding* é construído a partir dos conceitos existentes na ontologia. Conceitos relacionados ao desenho qualitativo do caso clínico, tais como, aspectos cognitivos, aspectos comportamentais ou hipótese diagnóstica são usados para ajuste na MHMO. Os conceitos da ontologia utilizados para o *coding* são criados na ferramenta Atlas-Ti (Figura 1a). Os casos clínicos são analisados qualitativamente. O *coding* é extraído com os dados codificados.

5. O framework FLAIR é executado em cada *coding* extraído a fim de reconhecer elementos textuais, tais como, sujeito, ação, complemento da ação, doença e informações adicionais dos tratamentos, a fim evidenciar as ações terapêuticas.

6. Os dados da estratificação de riscos e os elementos textuais reconhecidos na etapa anterior são organizados em seus respectivos *datasets* (Figura 1b e Figura 1c).

7. Anotação Semântica. A Tabela 1 traz o *Codebook*, que descreve os dados categoriais do *dataset*: *QualitativeAnalysis*,

⁶ <https://bioportal.bioontology.org/ontologies/MHMO>

⁷ <https://atlasti.com/>

RiskLevel e *Gender*. O DM (Tabelas 2 e 3) mapeia para ontologias (Sio e MHMO) as seguintes características para tratamento de saúde mental: *Patient*, *Name*, *Subject*, *Action*, *ComplementOfActon*, *Disease* e *AdditionalInformation*. Os grafos RDFs representam os fragmentos de conhecimento dos tratamentos psicoterápicos (Figura 1d).

5. Considerações Finais

O processo apresentado neste trabalho é fundamentado em modelagem ontológica e estabelece etapas para compreensão (com uso de ontologias) e preparação de dados para análise quali-quantitativa. Discute-se a aplicação do processo em dados da estratificação de risco em ansiedade e tratamentos psicoterápicos.

O trabalho utiliza ontologias para enriquecer o conhecimento, com o objetivo de auxiliar na compreensão dos dados. O processo é independente de uma ontologia específica, tanto para análise qualitativa, quanto para anotação semântica, indo além dos trabalhos correlatos citados (HAN e STOLFFEL, 2011; HOCKER *et al.*, 2021). A preparação de dados, apoiada por SDD, reutiliza o conhecimento disponível na ontologia a fim mapear os dados por meio de templates de metadados.

Argumentou-se neste trabalho que a geração de grafos de conhecimento, com o uso da anotação semântica pelo SDD, pode abrir caminho para análise de dados quantitativos e qualitativos sobre eficácia dos tratamentos, ou quais fatores que influenciam no tratamento.

Como trabalhos futuros, espera-se aplicar o processo apresentado utilizando a ferramenta HaDataAc⁸, visando a ingestão dos dados em uma infraestrutura que permita aquisições combinadas de dados e metadados semânticos. O uso do HaDataAc poderá ampliar a possibilidade da análise quali-quantitativa dos tratamentos em ansiedade a partir de variáveis harmonizadas em repositórios de dados semânticos.

Referências

AHIMA. Healthcare Data Governance . 2022. Disponível em: <https://rb.gy/pcjbdh>. Acesso em 03 de junho de 2022.

⁸ <https://www.hadatac.org/>

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019, June). FLAIR: An easy-to-use framework for state-of-the-art NLP. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations) (pp. 54-59).

BRISSEON, Laurent; COLLARD, Martine. An ontology driven data mining process. In: International Conference on Enterprise Information Systems. 2008. p. 54-61.

BERTHOLD, M.R., Borgelt, C., Hoppner, F., Klawonn, F.: Guide to intelligent data analysis: how to intelligently make sense of real data. Springer (2010).

BRASIL. Ministério da Saúde. Departamento de Atenção Básica. Cadernos de Atenção Básica: Núcleo de Apoio à Saúde da Família - Volume 1: ferramentas para a gestão e para o trabalho cotidiano. Brasília: Ministério da Saúde; 2014. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/nucleo_apoio_saude_familia_cab39.pdf

BRENAS, Jon Hael; SHIN, Eun Kyong; SHABAN-NEJAD, Arash. Adverse childhood experiences ontology for mental health surveillance, research, and evaluation: advanced knowledge representation and semantic web techniques. JMIR mental health, v. 6, n. 5, p. e13498, 2019.

CEUSTERS, Werner; SMITH, Barry. Foundations for a realist ontology of mental disease. Journal of biomedical semantics, v. 1, n. 1, p. 1-23, 2010.

HASTINGS, Janna *et al.* Representing mental functioning: Ontologies for mental health and disease. 2012.

HEALTHIT. Advancing Health Data and Metadata Standards. 2020. Disponível em: <https://www.healthit.gov/topic/scientific-initiatives/advancing-health-data-and-metadata-standards>. Acesso em 03 de junho de 2022.

HAN, Dong, and STOLFFEL, Kilianl. "Ontology based qualitative case studies for

sustainability research." In Proceedings of the AI for an Intelligent Planet, pp. 1-8. 2011.

HOCKER, J., BIPAT, T., MCDONALD, D. W., & Zachry, M. (2021). Developing an Ontology for Qualitative Coding Schemas - QualiCO. Disponível em: <https://eera-ecer.de/ecer-programmes/conference/26/contribution/50889/>. Acesso em 28 de jun de 2022.

KAPISZEWSKI, D., & KARCHER, S. (2021). Transparency in practice in qualitative research. *PS: Political Science & Politics*, 54(2), 285-291.

MELLO, Veronica de Pádua. Caminhos da educação em saúde na atenção básica: proposta de reorganização do grupo do parque. 2015. Tese de Doutorado. Universidade de São Paulo.

PAULA, George Luiz Costa de. Classificação de risco em saúde mental: implicações clínicas, éticas e sócio-políticas. 2019.

PARANÁ (Estado). Rede de Saúde Mental. Disponível em: <https://www.saude.pr.gov.br/Pagina/Saude-Mental>. Acesso em 09 de abr de 2021.

PATON, Norman. Automating data preparation: Can we? should we? must we?. In: 21st International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data. 2019.

PYLE, Dorian. Data preparation for data mining. Morgan Kaufmann, 1999.

RASHID, Sabbir M. et al. The Semantic Data Dictionary Approach to Data Annotation & Integration. In: SemSci@ ISWC. 2017. p. 47-54.

SCHRÖER, C., KRUSE, F., & GÓMEZ, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.

SVÁTEK, Vojtěch; RAUCH, Jan; RALBOVSKÝ, Martin. Ontology-enhanced association mining. In: *Semantics, Web and mining*. Springer, Berlin, Heidelberg, 2005. p. 163-179.

WHITEFORD, Harvey; FERRARI, Alize; DEGENHARDT, Louisa. Global burden of disease studies: implications for mental and substance use disorders. *Health Affairs*, v. 35, n. 6, p. 1114-1120, 2016.

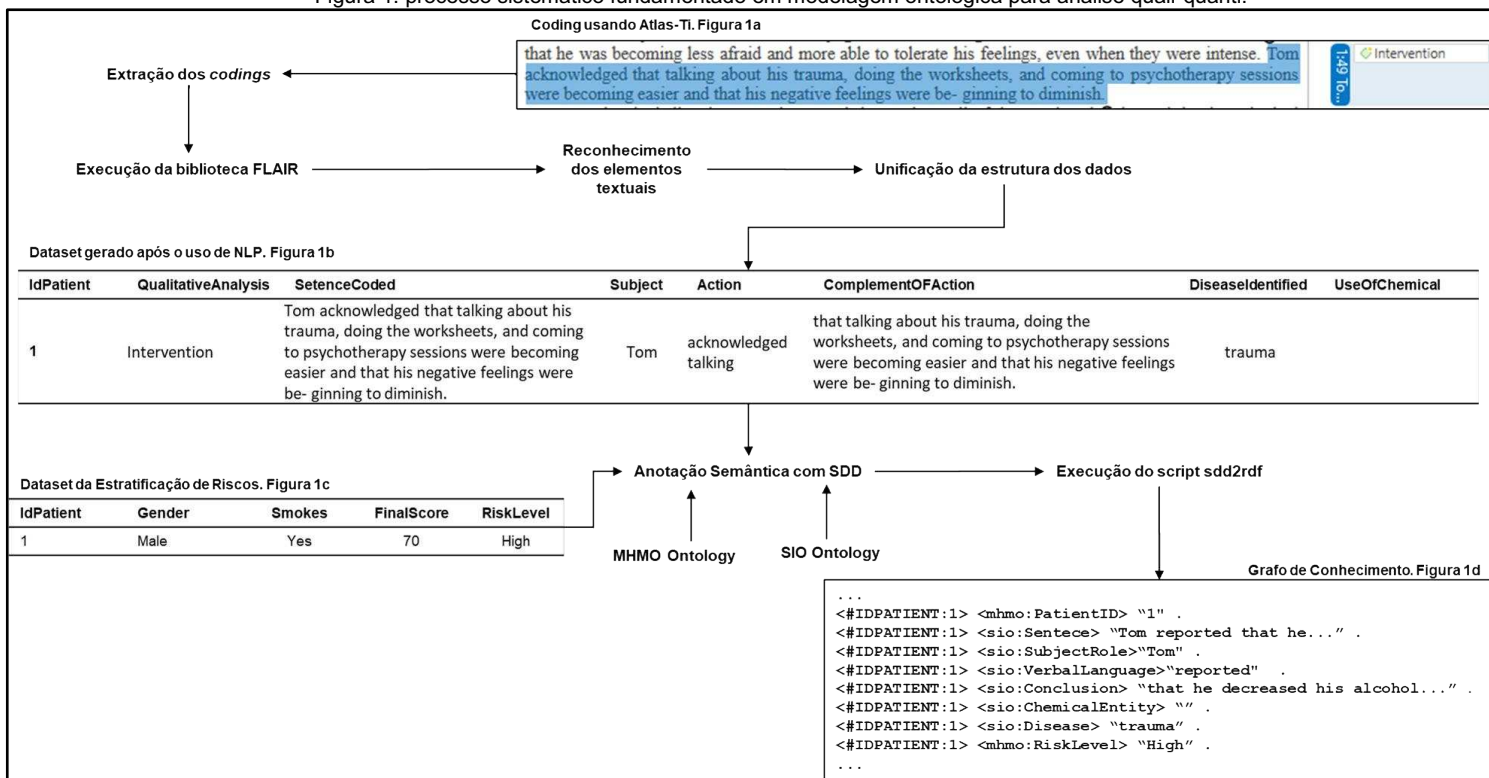
WILLIAMS, M., & MOSER, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, 15(1), 45-55.

YAMADA, Diego Bettiol *et al.* Ontology-Based Inference for Supporting Clinical Decisions in Mental Health. In: *International Conference on Computational Science*. Springer, Cham, 2020. p. 363-375.

ZAYAS-CABÁN, Teresa, and Jonathan S. Wald. "Opportunities for the use of health information technology to support research." *JAMIA open* 3, no. 3 (2020): 321-325.

Apêndice A

Figura 1. processo sistemático fundamentado em modelagem ontológica para análise quali-quantitativa.



Fonte: Elaborada pelos autores.

Tabela 1. Codebook das *.QualitativeAnalysis*, *RiskLevel* e *Gender*.

Column	Code	Label	Class
Gender	Male	Gender of patient	mhmo:PatientGender
Gender	Female	Gender of patient	mhmo:PatientGender
Smokes	Yes	Patient smokes	mhmo:Smoking
Smokes	No	Patient smokes	mhmo:Smoking
RiskLevel	High	Risk level stratified	mhmo: RiskLevel
RiskLevel	Medium	Risk level stratified	mhmo: RiskLevel
RiskLevel	Low	Risk level stratified	mhmo: RiskLevel
UseOfChemicalSubstance	alcohol	Chemical Substance	mhmo:AlcoholDependence
Disease	anxiety	Mental Disorder	mhmo:AnxietyDisorder
QualitativeAnalysisOfClinicalCase	Diagnosis	Qualitative Analysis	mhmo:Diagnosis
QualitativeAnalysisOfClinicalCase	UseOfSubstance	Qualitative Analysis	mhmo:SubstanceDependence

Fonte: Elaborada pelos autores.

Tabela 2. Especificação do DM para dados explícitos.

Column	Attribute	AttributeOf	Unit
IDPatient	mhmo:PatientID	??patient	schema:Integer
SentenceCoded	sio:Sentence	??treatment	schema:Text
Subject	sio:SubjectRole	??treatment	schema:Text
Action	sio:VerbalLanguage	??treatment	schema:Text
ComplementOfAction	sio:Conclusion	??treatment	schema:Text
DiseaseIdentified	sio:Disease	??patient	schema:Text
UseOfChemical	sio:ChemicalEntity	??chemicalsubstance	schema:Text

Fonte: Elaborada pelos autores.

Tabela 3. Especificação do DM para dados implícitos.

Column	Entity	Relation	InRelationTo
??patient	sio:Human	mhmo:receives	??treatment
??patient	sio:Human	mhmo:IsDiagnosedWith	mhmo:MentalDisorder
??patient	sio:Human	mhmo:takes	??chemicalsubstance
??patient	sio:Human	mhmo:hasRiskStratifiedIn	??risklevel
??patient	sio:Human	mhmo:uses	??chemicalsubstance

Fonte: Elaborada pelos autores.