

Ontologia para a Prevenção de Crimes Sexuais Online Contra Crianças e Adolescentes

Denise F. Tsunoda¹, José M. A. P. Cestari¹, Alex S. Constâncio² and Luciano H. G. Marin¹
¹Federal University of Paraná, Av. Prefeito Lothario Meissner, 632 – Jardim Botânico, Curitiba, 80210-170, Brazil

² Federal University of Paraná, Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba - PR, 81530-000, Brazil

Abstract

The Child and Adolescent Statute and the General Data Protection Law brought important advances in Brazil regarding the prevention and protection of children and adolescents against various types of crimes. These laws, as well as others that protect children and adolescents against crimes of sexual violence, require constant updates when considering the ease of access and use of information and communication technologies by all individuals in our society. As far as our research can be carried out, however, there is no systematization of the theme of sexual violence against children and adolescents in relation to its extensive number of related terms, increasing the chances of inappropriate use of terminologies and consequent misinformation, weakening initiatives to combat it. Faced with the challenge of systematizing the theme of sexual violence against children and adolescents, this article describes the creation of an ontology with the terms selected for this theme, the methodology adopted for its creation and the criteria for acceptance and revision of the chosen terms, as well as the definition of classes, subclasses, instances and relationships. Potential uses of the resulting ontology are also presented, aiming to facilitate access to terms and their respective laws by society in general.

Keywords

Ontology; Sexual Violence; Children and Adolescents.

Resumo

O Estatuto da Criança e do Adolescente e a Lei Geral de Proteção de Dados trouxeram importantes avanços no Brasil em relação à prevenção e a proteção de crianças e adolescentes contra diversos tipos de crimes. Essas leis, assim como outras que protegem crianças e adolescentes contra crimes de violência sexual, requerem atualizações constantes quando se considera a facilidade de acesso e uso das tecnologias de informação e comunicação por todos os indivíduos de nossa sociedade. Até onde nossas pesquisas puderam ser realizadas, no entanto, não há uma sistematização do tema violência sexual contra crianças e adolescentes em relação ao seu extenso número de termos relacionados, aumentando as chances de uso inadequado de terminologias e consequente desinformação, fragilizando as iniciativas de combate a este tema. Diante do desafio de sistematizar o tema violência sexual contra crianças e adolescentes, este artigo descreve a criação de uma ontologia com termos selecionados para esse tema, a metodologia adotada para sua criação e os critérios de aceitação e a revisão dos termos escolhidos, bem como a definição de classes, subclasses, instâncias e relacionamentos. Também são apresentados potenciais usos da ontologia resultante, visando facilitar o acesso aos termos e suas respectivas legislações pela sociedade em geral.

Palavras-chave

Ontologia; Violência Sexual; Criança e Adolescente.

1. Introdução

Proceedings of the XVI Seminar on Ontology Research in Brazil (ONTOBRAS 2023) and VII Doctoral and Masters Consortium on Ontologies (WTDO 2023), Brasilia, Brazil, August 28 - September 01, 2023.

EMAIL: dtsunoda@ufpr.br (A. 1); cestari@ufpr.br (A. 2); alex.constancio@ufpr.br (A. 3); luciano.gallegos@ufpr.br (A. 4)

ORCID: 0000-0002-5663-4534 (A. 1); 0000-0002-4331-6588 (A. 4)

© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

A legislação brasileira, por meio da Lei 8.069 de 13 de julho de 1990 - Estatuto da Criança e do Adolescente/ECA [1], dispõe sobre os direitos de crianças e adolescentes no Brasil. Após a criação do ECA, outras leis relacionadas surgiram, a exemplo da Lei 12.965 de 23 de abril de 2014 - Marco Civil da Internet, e Lei nº 13709 de 14 de agosto de 2018 - Lei Geral de Proteção de Dados/Nova LGPD [2]. Ainda assim, como qualquer legislação, estas leis requerem atualizações constantes, principalmente quando considerados a onipresença, a facilidade de acesso e o uso das tecnologias de informação e comunicação por todos os indivíduos da sociedade, incluindo crianças e adolescentes. A acessibilidade a termos destas leis pode auxiliar no combate à violência sexual contra crianças e adolescentes.

Os termos envolvendo a temática sobre violência sexual contra crianças e adolescentes encerra um conjunto extenso de termos relacionados que, até onde se pôde levantar, não foi ainda sistematizado por meio de uma representação de conhecimentos não apenas para o Brasil, mas em escala mundial. A ausência deste tipo de representação, tal como ocorre em ontologias, para este tema aumenta as chances de uso inadequado de terminologia e consequente desinformação, que pode enfraquecer as iniciativas para o seu combate. Termo, para efeitos desta pesquisa, é qualquer palavra ou expressão de interesse que seja utilizada e conotada dentro do domínio em estudo, neste caso, a violência sexual online contra crianças e adolescentes. Exemplos de termos incluem “abandono”, “chat”, “online threats”, “organização criminosa” e “abandono de incapaz”.

Neste trabalho, como pesquisa e como questão disciplinar, considera-se que uma ontologia (tipo de representação de conhecimentos) tem foco orientado ao fornecimento de sistemas de categorização para organizar a realidade [3] ou uma especificação explícita de uma conceitualização [4], estendendo-se também como “especificação explícita e formal dos termos de um domínio e das relações entre eles” . Studer et al. [5] mencionam que uma ontologia é uma “especificação formal e explícita de uma conceitualização compartilhada”. A representação de uma ontologia demanda um formalismo, e Maedche et al. [6] propuseram o formalismo de 5-tuplas, considerando conceitos, relacionamentos, hierarquia de relacionamentos, conjuntos de classes e um conjunto de axiomas. A metodologia 101 [7] sugere as seguintes etapas para a construção de uma ontologia: a) determinar domínio e escopo da ontologia; b) considerar o reuso de outras ontologias; c) enumerar os termos importantes da ontologia; d) definir classes e hierarquia de classes; e) definir propriedades de classes; f) definir os valores de propriedades; e g) criar instâncias. De qualquer forma, uma das etapas para a construção de qualquer ontologia é a definição dos termos que farão parte dela.

A compreensão de um tema de interesse, e dos termos mais relevantes deste tema, depende de conhecimento a respeito das conotações particulares que palavras e expressões assumem naquele domínio. Igualmente, a elaboração de quaisquer conteúdos escritos, notadamente normas jurídicas, exige um vocabulário consistente que enumere, conceitue e relacione termos e expressões, para propiciar um entendimento comum entre as partes interessadas. Uma ontologia fornece, além de um vocabulário, construções lógicas e semânticas específicas de um determinado campo, tornando-a um artefato ideal para a conceituação e difusão clara de um tema.

A principal contribuição desta pesquisa é a elaboração da "Child-Safe", ontologia que reúne, organiza e relaciona as terminologias operadas no domínio da prevenção de crimes sexuais contra crianças e adolescentes. O propósito desta ontologia foi o de estabelecer o vocabulário comum para o tema, além das construções lógicas e semânticas que relacionam os conceitos.

A Child-Safe foi desenvolvida em atendimento aos requisitos estabelecidos em um edital de chamamento público do Ministério da Mulher, da Família e dos Direitos Humanos (atual Ministério dos Direitos Humanos e Cidadania) nos termos do Projeto de Execução Nacional BRA/10/007 “Boas práticas na implantação e implementação dos sistemas de informação para a infância e a adolescência - SIPIA, Conselhos tutelares e SIPIA, SINASE WEB”, nos anos de 2021 e 2022. O edital previa apenas a concepção da ontologia em curto prazo para entrega, uma vez que os procedimentos de validação, avaliação e uso estariam contemplados em editais subsequentes.

Diante do desafio em produzir uma ontologia que identifique e conceitue os aspectos de interesse a respeito do tema, neste artigo descreve-se também a metodologia adotada,

pormenorizando como os termos foram selecionados, os critérios adotados para o seu cadastro, aceitação e revisão, além da definição de classes, subclasses, instâncias e relacionamentos. São também apresentados potenciais usos da ontologia resultante, concebida para prevenção de crimes sexuais online contra crianças e adolescentes, visando facilitar o acesso aos termos e suas respectivas leis pela sociedade em geral.

O presente artigo está organizado em seções, conforme segue: a seção “Metodologia de Elaboração da Ontologia” descreve as etapas, recursos e processos que viabilizaram a construção da ontologia; a seção “Dicionário Semântico” descreve os passos para a inclusão de termos, definições e relacionamento de palavras envolvendo crimes cibernéticos aplicados a crianças e adolescentes; a seção “Ontologia Resultante” explica a exportação do dicionário semântico para uma ferramenta de ontologias visando o enriquecimento semântico e a seção “banco de dados semântico” descreve o armazenamento e a visualização da ontologia resultante em um banco de dados não-estruturado na forma de grafo. A seção “uso da ontologia proposta” pormenoriza as possíveis utilizações da ontologia resultante e de sua visualização por meio de banco de dados semântico e, finalmente, na seção “considerações finais” são apresentadas as impressões finais deste artigo.

2. Metodologia de Elaboração da Ontologia

O processo para a elaboração da ontologia sobre violência sexual contra crianças e adolescentes foi dividido em duas grandes etapas: o *dicionário semântico* contendo todo o corpo de vocabulário que pôde ser encontrado e organizado a respeito do tema; e a *ontologia*, derivada do dicionário semântico e contemplando relacionamentos entre os termos convertidos em classes.

O *dicionário semântico* foi mantido em uma planilha compartilhada em nuvem, organizada para descrever detalhadamente cada um dos termos elicitados. Este enriquecimento prevê não apenas a descrição do termo propriamente por meio de diversos metadados, mas também prepará-lo para servir de insumo para a construção da ontologia.

Um processo de revisão foi estabelecido, visando reduzir a chance de vieses e vícios. Para tanto, cada termo do dicionário contou com a identificação de seu autor (o membro que introduziu o termo no dicionário) e dos seus revisores. Revisores de texto e de semântica foram destacados, tanto para garantir a qualidade da redação dos descritivos de cada termo, como também de sua consistência jurídica e semântica, dentro o tema próprio da ontologia.

A *ontologia* resultante foi derivada do *dicionário semântico* por meio de uma ferramenta escrita em linguagem de programação Python (disponível em repositório público no GitHub³), que converteu o arquivo em planilha para um arquivo fonte em convenção Turtle (.TTL). O formato TTL foi escolhido por ser de mais fácil interpretação e operação por seres humanos, quando comparado com o formato OWL (que é uma especialização do formato XML). Outro motivo foi por ser compatível com o editor de ontologias escolhido, o Protégé e sua versão online, o WebProtégé. A ontologia possui termos e metadados definidos do dicionário semântico e inclui os relacionamentos semânticos entre as diversas classes concebidas. Nas seções seguintes, os pormenores do dicionário semântico e da elaboração da ontologia serão descritos.

3. Dicionário Semântico

A construção de uma ontologia depende da definição e do relacionamento dos termos que a constituirão. Neste trabalho, a seleção dos termos relativos a crimes sexuais online aplicados a crianças e adolescentes foi realizada por meio de pesquisa em base de dados de periódicos científicos e de fontes brasileiras de informações, tais como: legislações (ex.: Constituição da República Federativa do Brasil [9], Estatuto da Criança e do Adolescente [1], Código Penal [10] e

³ <https://github.com/gambit4348/BRA-10-007-Produtor>

Lei Geral de Proteção de Dados Pessoais [2]), cartilhas, manuais e outros documentos relacionados em organizações governamentais e não governamentais (ex.: Comitê Gestor da Internet [11], Safernet [12] e TIC Kids Brasil [13]), e Classificação Internacional de Doenças (CID) [14]. Foram, também, utilizadas fontes internacionais, como da Interpol e sua *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse* [15], atualizado em 2016, onde há uma listagem de acrônimos (p. 103-104) que foi filtrada e complementada com outros termos recuperados na internet e incluídos no dicionário.

O processo de eliciação de termos que vieram a compor o dicionário foi semiautomatizado. O processo automatizado de extração de termos é um processo bem estabelecido pela comunidade científica e tem o objetivo de acelerar o processo de extração de termos de forma computadorizada e estabelecida nos moldes do processamento de linguagem natural, onde termos relevantes são selecionados e suas variações semânticas são inclusas, de forma a melhorar a composição do dicionário semântico [16]. O processo automatizado de extração de termos pode ser melhorado com a utilização de boas práticas de extração de terminologias dentro de um determinado escopo temático, levando-se em consideração uma padronização na extração de tais termos baseados em ano, país e objetivo [17].

Considerando o processo automático de extração de termos, e suas boas práticas, na etapa automatizada fez-se o uso do resultado de buscas em bases de dados de periódicos. O corpus de documentos foi posteriormente submetido ao Biblioshiny [18], utilizado para extrair diversos termos relacionados. Na etapa manual, o dicionário foi complementado por especialistas em diversas áreas (Direito, Tecnologia da Informação, Gestão da Informação) com termos oriundos dos diversos documentos citados, extraindo e, eventualmente, combinando definições, seus significados e seus metadados.

A pesquisa em bases de dados de periódicos científicos foi fundamental para identificar os termos utilizados pelos pesquisadores como palavras-chave. Devido ao caráter interdisciplinar do trabalho para o desenvolvimento da ontologia descrita neste artigo, optou-se por pesquisar bases de dados multidisciplinares. Foram selecionadas as bases de dados Web of Science, Scopus e Dimensions, em 22 de setembro de 2021, com a seguinte estratégia de busca: (cybercrime OR cybersecurity OR cyberstalking OR sextortion OR "revenge porn" OR "online threats" OR "Child Sexual Abuse Material" OR "Child Sexual Exploitation Material") AND (child OR minor OR teenager OR adolescent OR juvenile OR stripling OR sapling OR "young women" OR "young woman" OR "young boy"), do qual foram recuperados 517 documentos, identificadas 1.367 palavras-chaves únicas de autores e 2.002 palavras-chave estendidas.

As análises preliminares sobre os documentos consideraram os termos com ocorrência igual e superior a 10 no conjunto de documentos recuperados, resultando em: (a) 80 palavras-chaves, (b) 12 termos recuperados dos títulos pela análise por unigramas, (c) 14 termos recuperados dos títulos pela análise por bigramas e trigramas, (d) 84 termos recuperados dos resumos pela análise por bigramas e (e) 10 termos recuperados dos resumos pela análise por trigramas. Estes resultados foram reunidos, os termos duplicados foram retirados após a tokenização manual, e 75 termos restaram e foram traduzidos para a língua portuguesa e inseridos em um dicionário (ex.: "child exploitation" foi inserido no dicionário como "exploração infantil"). O cadastro e validação do conjunto de termos foi realizado por meio de uma planilha eletrônica em nuvem e colaborativa, e a cada termo foi atribuído o seguinte conjunto de metadados (colunas) na planilha:

1. Termo: campo do tipo alfanumérico, obrigatório, de entrada do termo no dicionário semântico. Exemplo: "chat";
2. Relacionamento: coluna opcional, do tipo alfanumérica que identifica os relacionamentos existentes entre o (1) Termo e outros termos. Exemplo: "evitado por" como relacionamento entre "violação de sigilo processual" (coluna Termo) e "cibersegurança" (coluna Termo relacionado);
3. Termo relacionado: coluna opcional, do tipo alfanumérica que identifica o(s) termo(s) relacionado(s) ao (1) Termo pelo(s) (2) Relacionamento(s) definidos (coluna Relacionamento);
4. Superclasse: coluna opcional, do tipo alfanumérica que identifica os termos gerais de um termo principal. Exemplo: "Agressão" e "Violência" são superclasses de "ciberviolência";

5. Exemplo de: coluna opcional, do tipo alfanumérica, que identifica quando um termo é um caso particular de uma coleção. Ontologicamente, identifica a instância de uma classe. Exemplo: “Homofobia” (Termo) é um tipo de “Crime” (Classe);
6. Equivalente: campo do tipo alfanumérico, opcional, que identifica termo(s) sinônimos(s) do (1) Termo principal. Exemplo: “bate-papo” como termo equivalente de “chat”;
7. Descrição: campo alfanumérico, obrigatório, apresenta a descrição do termo cadastrado. Exemplo: “O abuso infantil é o uso da própria força, de um mau, excessivo injusto contra crianças e adolescentes.”;
8. Eixo: campo alfanumérico, obrigatório, que identifica a(s) área(s) do conhecimento ou de aplicação que originam ou utilizam o termo. Exemplo: “Aborto” é pertencentes aos eixos “Direito” e “Geral”, pois o Direito o pode conotar como atividade ilegal ao mesmo tempo que também é um termo de uso geral.
9. Terminologia recomendada: campo binário, admitindo “Sim” ou “Não”, obrigatório, que classifica o termo como de uso recomendado ou não. Exemplo: “Menor” é um termo de uso corriqueiro que tem tido sua utilização desestimulada em função da forte carga de negatividade que historicamente acumulou.
10. Fonte: campo alfanumérico, obrigatório, que identifica as fontes literárias onde o termo foi identificado, eventualmente cunhado ou caracterizado. Exemplo: tanto o Código Penal Brasileiro de 1940, como a Enciclopédia Jurídica PUC-SP apresentam definições para “Aborto induzido ilegal”.
11. Link: campo alfanumérico, opcional, para cadastro do link no qual a fonte utilizada está disponível. Exemplo: https://www.ohchr.org/Documents/Issues/Children/SR/TerminologyGuidelines_en.pdf;
12. Criador: campo alfanumérico obrigatório, com a identificação da pessoa responsável pelo cadastro do termo no dicionário. Exemplo: Aloysio;
13. Revisor: campo alfanumérico obrigatório, com a identificação da pessoa responsável pela revisão técnica do termo e descrição. Exemplo: Marcelo;
14. Revisor textual: coluna de preenchimento obrigatório, do tipo alfanumérica que identifica o revisor textual da coluna Descrição. Exemplo: José da Silva;

Na estrutura da planilha, contendo 14 colunas e acima apresentada, torna-se possível constatar a preocupação com a concepção da ontologia, uma vez que elementos tais como classe, superclasse e relacionamento serão destacados com este propósito, a partir do (1) Termo principal e de suas demais características. No entanto, as classes e os relacionamentos foram incluídos somente após a finalização do dicionário, quando novos termos não estavam mais sendo incluídos.

Após as etapas de criação e validação, o dicionário semântico passou a contar com 506 termos principais cadastrados. A ferramenta (desenvolvida em linguagem Python) converteu a planilha para um arquivo HTML estático com código em Javascript, artefato que passou a ser chamado de “Navegador do Dicionário”. O conversor de planilha mapeia os metadados do dicionário semântico para construções ontológicas, e será pormenorizadamente descrito na próxima seção.

O Navegador do Dicionário pode ser carregado em um *browser*, oferecendo melhor visualização e facilidades como ordenação alfabética e hiperlinks. Este pode ser considerado o primeiro uso do dicionário. Uma imagem desta conversão, com o seu Navegador, pode ser vista na **Figura 1**, onde o dicionário semântico demonstra os termos que podem ser separados por eixos (Direito, Criança e Adolescente, Tecnologia, Saúde e Geral) ou ainda filtrados pela letra inicial.

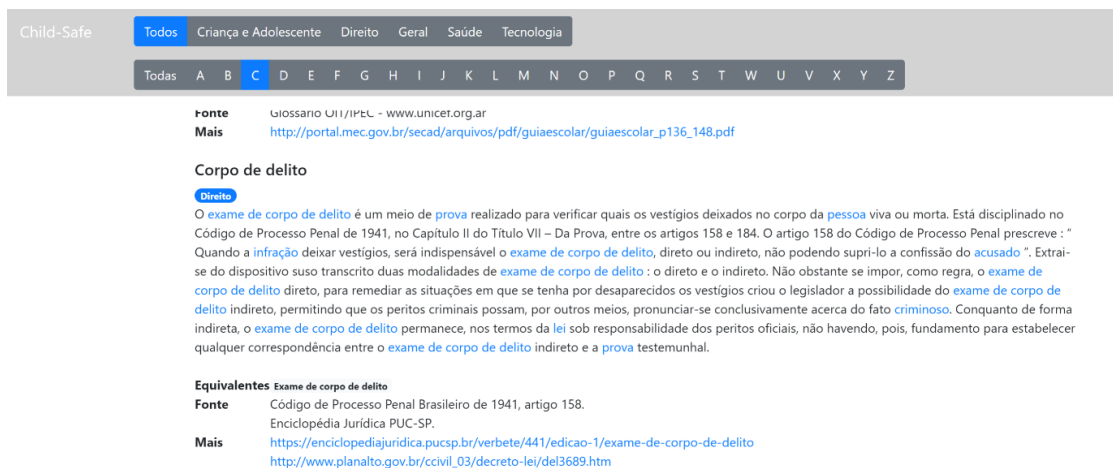


Figura 1: Navegador do Dicionário, em formato HTML, que apresenta os principais metadados no qual um termo pode ser categorizado e separados alfabeticamente para facilidade de operação. Demais expressões que fazem parte do dicionário semântico são mostradas como hiperlinks, para facilitar a navegação dentro do vocabulário.

Alguns termos podem estar atribuídos a mais de um eixo de conhecimento humano, a exemplo de “Menor” que, inclusive, é um termo classificado como não recomendado, conforme apresentado na **Figura 2**. Os eixos constituem as categorias nas quais um termo pode estar vinculado, sendo frequentemente utilizado naquele domínio. Os termos “não recomendados” (em vermelho) são aqueles que fazem parte do vocabulário do tema, mas que por motivos históricos assumiram conotações negativas e/ou ofensivas.

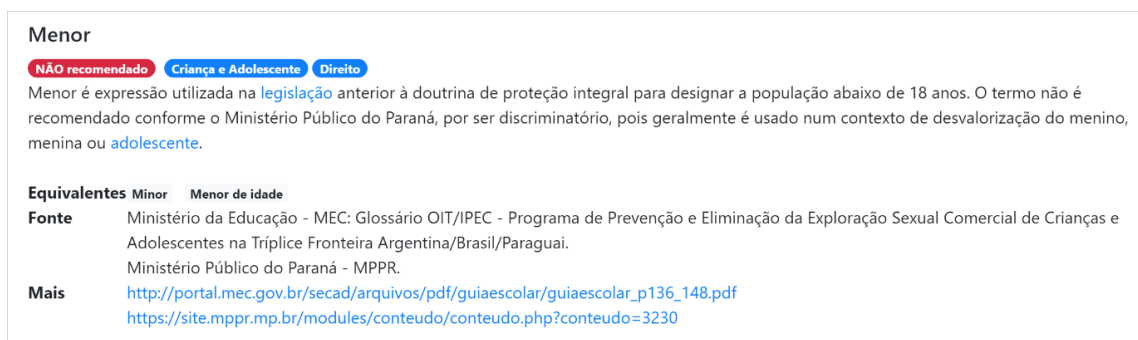


Figura 2: Detalhe do termo “Menor”, que pertence aos eixos “Criança e Adolescente” e “Direito”, marcado como “Não recomendado”; as fontes literárias que originaram o termo estão mostradas no campo “Fonte” e links para os documentos relacionados são mostrados no campo “Mais”.

Apesar de ser um arquivo HTML local e não depender de acesso à internet, o Navegador do dicionário semântico é funcional e dinâmico, com todos os seus links e menus, ativos. Atualizações e correções feitas na planilha inicial do dicionário semântico são facilmente convertidas para o Navegador por meio do conversor de planilhas. A configuração final do dicionário semântico, devidamente convertido em formato Turtle, foi exportada para as ferramentas de ontologias Protégé, WebProtégé, e mais tarde para o banco de dados semântico e em grafo Neo4j, a ser detalhado nas próximas duas seções.

4. Ontologia Resultante

Nos primeiros momentos da elaboração da ontologia, foi experimentado o uso de uma ontologia de topo chamada gUFO [19], que é, como se autodefine, uma versão *lightweight* da UFO (Unified Foundational Ontology). A gUFO oferece um conjunto genérico de construções semânticas para

abstrações variadas como tempo, espaço, eventos, indivíduos, entre outros, e está livremente disponível. Embora o Protégé desktop interpretasse corretamente a diretiva de importação do documento gufo.ttl, o mesmo não se deu com o WebProtégé. Depois de diversas tentativas de se utilizar a gUFO como ontologia de topo para a Child-Safe, e atendendo aos prazos estabelecidos no edital de chamamento público, optou-se por construir a ontologia sem o auxílio de uma ontologia de topo naquele momento.

O conversor de planilha, além da carga de dados no Navegador, permite o uso do dicionário semântico em uma ontologia, por meio de um documento em formato Turtle (Terse RDF Triple Language), uma convenção sintática para a escrita textual de triplas RDF (Resource Description Framework) que podem ser utilizadas como blocos fundamentais para a expressão de ontologias digitais. O Turtle é um dos formatos prescritos pelo World Wide Web Consortium (OWL) [3, 20].

O arquivo de planilha (.XLSX) que contém o Dicionário semântico é utilizado como entrada e a saída é um documento em texto plano no formato Turtle (.TTL) que descreve a ontologia correspondente. Na **Tabela 1** está apresentado o mapeamento de cada metadado presente no dicionário semântico para a respectiva construção ontológica, em sintaxe Turtle.

Tabela 1: Mapeamento entre metadados do Dicionário e respectivas construção ontológica e construção Turtle convencionadas pelo Produtor.

Metadado	Construção ontológica	Construção Turtle
Termo	Classe, que identifica um conceito de interesse, representada por {name}	:{name} rdf:type owl:Class
Relacionamento e Termo relacionado	Propriedade, que identifica o relacionamento, representado por {relacionamento}, com outra classe, representada por {class.name}	:{relacionamento} :{class.name}
Superclasse	Superclasse, que identifica um conceito mais genérico do Termo, representado por {superclasse}	rdfs:subClassOf: {superclasse}
Exemplo de	Indivíduo, que identifica um caso concreto da classe do termo, representado por {name}	:{name} rdf:type owl:NamedIndividual
Equivalente	Classe equivalente, que identifica uma outra classe que apresenta os mesmos indivíduos da classe do termo, representada por {equivalente.name}	owl:sameAs :{equivalente.name}
Descrição	Comentário contendo o texto descritivo da classe, representado por {texto}	rdfs:comment "{text}"@pt-BR
Eixo	Domínio, que identifica os domínios relacionados com o Termo, representado por {eixo}	:definedIn "{eixo}"@pt-BR
Terminologia recomendada	Anotação que identifica se o termo é ("Sim") ou não ("Não") recomendada	:isRecommendedTerm "Sim"@pt-BR :isRecommendedTerm "Não"@pt-BR
Fonte	Anotação que identifica a fonte que originou o termo, representada por {fonte}	:sourceIs "{fonte}"@pt-BR
Link	Anotação que informa um link que contém um documento relacionado ao termo, representado por {link}	:linkIs "{link}"
Criador	Não convertido	
Revisor	Não convertido	
Revisor textual	Não convertido	

Algumas convenções foram estabelecidas para assegurar a consistência do processo de geração da ontologia: (a) os termos listados no dicionário semântico geram classes na ontologia, ainda sem a formação de taxonomias; (b) os metadados descritivos (Definição, Eixos, Fontes, Links e Termo Recomendado) do termo geram anotações na classe correspondente; (c) termos equivalentes geram outras classes vinculadas, referenciadas por meio de como "sameAs", que estabelece que as duas classes são sinônimas, pois apresentam os mesmos indivíduos; e (d) relacionamentos e termos relacionados geram propriedades de classe, criando ligações semânticas entre as classes envolvidas.

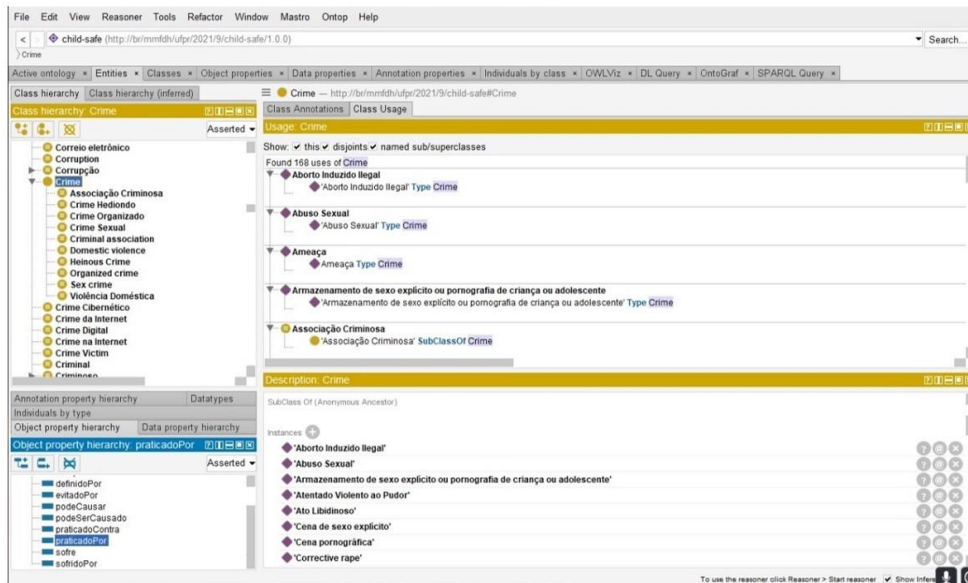


Figura 3: Classe “Crime”, da ontologia resultante, mostrada na ferramenta Protégé, com destaque para subclasses (ex.: “Associação Criminosa” e “Crime Hediondo”), relacionamentos (ex.: “definidoPor” e “evitadoPor”) e referências (ex.: “Aborto Induzido Ilegal” e “Abuso Sexual”).

O produto da conversão do dicionário semântico conta com 506 termos principais que geraram 863 classes no Protégé, pois alguns termos equivalentes da planilha original foram convertidos para classes na ontologia. Como exemplo, a **Figura 3** demonstra alguns relacionamentos da classe “Crime”, que apresenta a subclasse “Associação Criminosa”, enquanto “Ameaça”, um de seus casos particulares. Um trecho do código Turtle que descreve a classe “Aborto Induzido Ilegal” é apresentado na **Figura 4**: Trecho de código Turtle apresentando a definição da classe “Aborto Induzido Ilegal”, subclasse de “Crime”, assim como seus relacionamento e anotações..

```

### http://br/mmfdh/ufpr/2022/2/child-safe#Aborto_Induzido_Ilegal
:Aborto_Induzido_Ilegal rdf:type owl:Class ;
  rdfs:subClassOf :Aborto_Induzido ;
  rdfs:subClassOf :Crime ;
  rdfs:subClassOf :Condição ;
  rdfs:label "Aborto induzido ilegal"@pt-BR ;
  :definedIn "Geral"@pt-BR ;
  :definedIn "Direito"@pt-BR ;
  :isRecommendedTerm "Sim"@pt-BR ;
  :sourceIs "Código Penal Brasileiro de 1940... Enciclopédia Jurídica PUC-SP."@pt-BR ;
  :linkIs "http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm" ;
  :linkIs "https://enciclopediajuridica.pucsp.br/verbete/410/edicao-1/aborto" ;
  rdfs:comment ""Aborto induzido ilegal, ...""@pt-BR ;
  owl:sameAs :Aborto_Provocado_Ilegal .

:Aborto_Induzido_Ilegal rdf:type owl:namedIndividual ;
  rdf:type :Aborto_Induzido_Ilegal ;
  rdfs:label "Aborto induzido ilegal"@pt-BR ;
  :facilitadoPor :Casa_De_Aborto .

```

Figura 4: Trecho de código Turtle apresentando a definição da classe “Aborto Induzido Ilegal”, subclasse de “Crime”, assim como seus relacionamento e anotações.

O arquivo de ontologia gerado foi carregado nas ferramentas Protégé e WebProtégé. A versão *desktop* do Protégé (**Figura 3**) foi utilizada porque oferece um motor de inferência (*reasoner*) que permite avaliar a consistência das relações semânticas expressas e desdobrar a relações implícitas entre as classes. O *reasoner* (neste caso, o Hermit) do Protégé foi utilizado para avaliar a consistência lógica dos diversos axiomas expressos no arquivo Turtle e serviu para retroalimentar o processo de elaboração dos relacionamentos entre termos do dicionário semântico. As indicações de inconsistência geradas pelo *reasoner* foram utilizadas para revisão e correção daqueles relacionamentos.

A ontologia, então, é constituída de todos os termos do dicionário semântico com as relações de hierarquia, instâncias e relacionamentos específicos do domínio. Cada classe gerada apresenta os atributos do vocabulário e identifica os termos equivalentes como classes, além dos relacionamentos. **Figura 5** Na **Figura 5** possível visualizar a classe “Crime” com seus diversos detalhes, incluindo os relacionamentos “praticadoPor” com as classes “Criminoso” e “Reincidente”.

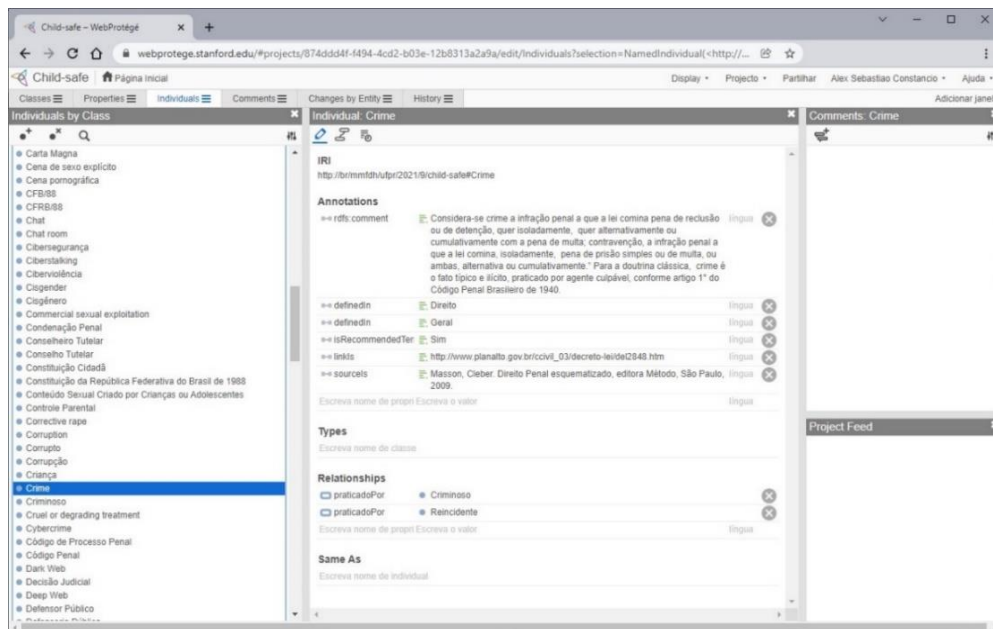


Figura 5: Classe “Crime”, na ferramenta WebProtégé, com destaque para as anotações (metadados oriundos do Dicionário) e relacionamentos com as classes “Criminoso” e “Reincidente”.

Embora o WebProtégé não possuísse um *reasoner* à época em que o seu uso se deu, ele foi utilizado para operação colaborativa entre membros da equipe em diversas etapas do desenvolvimento da ontologia.

A mesma classe “Crime” está representada na **Figura 3** e na **Figura 5**, mas as visualizações são ligeiramente distintas nas ferramentas Protégé e WebProtégé, dadas as peculiaridades de cada uma delas. Mesmo assim, ambas as ferramentas foram adotadas, uma vez que oferecem recursos complementares, e a mesma ontologia resultante pode ser utilizada em outras opções para armazenamento e visualização semântica em formato de grafos, como no caso dos bancos de dados semânticos.

5. Banco de Dados Semântico

Os bancos de dados semânticos, tais como aqueles organizados em grafos, permitem o armazenamento de ontologias e a realização de pesquisa integradas a sistemas de informação. Dentre os mais recentes bancos de dados em grafo deste tipo, destaca-se o Neo4j [21] um Sistema Gerenciador de Banco de Dados desenvolvido pela Neo4j Inc, tendo por característica a utilização de estruturas de grafos para armazenamento e consultas semânticas em nós, arestas e propriedades.

O Neo4j permite a utilização de ontologias de duas formas: por meio da interoperabilidade para os casos de ontologias e padrões compartilhados de ontologias, e por meio de inferências para a dedução de novos conhecimentos a partir do conhecimento existente e, desta forma, derivar um novo conhecimento a partir de informações pré-existentes, inclusive em formato Turtle. A importação e a exportação de uma ontologia para o Neo4j Desktop podem ser realizadas

por meio do plug-in Neosemantics [21, 22], que permite a importação da ontologia resultante sobre prevenção de crimes contra abuso de crianças e adolescentes descrita neste artigo.

Observações mais específicas da ontologia podem ser obtidas por meio do Neo4j Browser. Neste caso, pode-se escrever diretamente o nome da classe ou subclasse, ou relacionamento da ontologia por meio de comando Cypher, ou restringir a visualização da ontologia a alguns nós e relacionamentos, como pode ser observado na **Figura 5**. Neste exemplo, apenas 25 relacionamentos foram retornados por meio do comando Cypher "MATCH p=()->() RETURN p LIMIT 25".

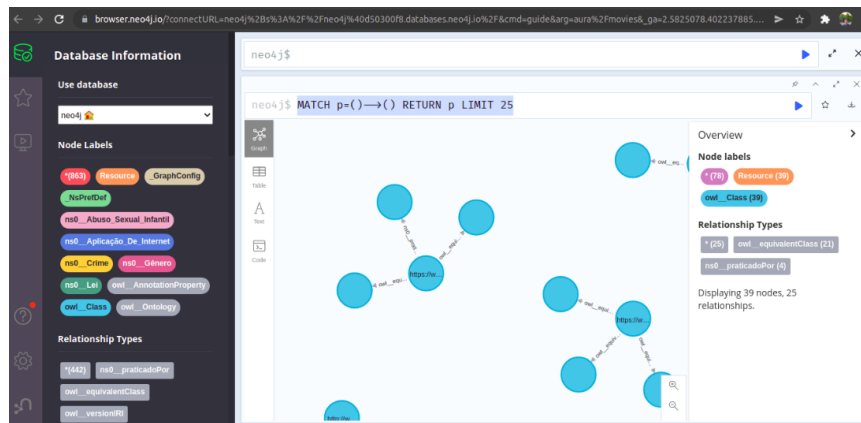


Figura 5: Visualização de 25 relacionamentos da ontologia retornados por meio do comando Cypher "MATCH p=()->() RETURN p LIMIT 25". Note que a visualização foi aumentada e, portanto, apenas uma parte dos 25 relacionamentos é visualizado.

6. Usos da ontologia proposta

Os autores Chandrasekaran, Josephson e Benjamins, na obra "What are ontologies, and why do we need them?" comentam sobre a utilidade das ontologias na inteligência artificial (IA), conforme traduzido e apresentado por [23] como duas grandes categorias para as teorias de IA: teorias de mecanismos e teorias de conteúdo. Ontologias são teorias de conteúdo sobre os tipos de objetos e as relações entre objetos que são possíveis em um domínio específico de conhecimento. Elas fornecem potenciais termos para descrever o conhecimento sobre o domínio. Algumas outras possíveis aplicações da ontologia, após compilação e adaptação ao contexto da pesquisa, são:

1. padronização e interoperabilidade: a padronização de terminologia proporcionada pelas ontologias permite que diferentes sistemas e instituições possam interoperar entre si, compartilhando informações e conhecimentos relevantes para a prevenção de crimes online contra crianças e adolescentes;
2. apoio à tomada de decisão: o desenvolvimento de sistemas de auxílio na adoção de medidas preventivas para proteção de crianças e adolescentes contra crimes online pode ser aperfeiçoado com o conhecimento estruturado das ontologias em situações tais como: configurações de privacidade, controle parental, fatores de riscos, estatísticas, boas práticas e outros;
3. identificação de riscos: a avaliação da possibilidade de um evento prejudicial para crianças e adolescentes na internet, tais como abuso e exploração sexual, pode ser facilitada com o uso de uma ontologia uma vez que essa define propriedades que descrevem características relevantes para a identificação desses riscos tais como o uso de linguagem e conteúdo impróprios;
4. monitoramento de mídias sociais: as ontologias podem ser usadas para monitoramento automático de mídias sociais (filtragem de conteúdo, detecção de eventos, análise de sentimentos e emoções, dentre outros), identificação e denúncia de comportamentos suspeitos ou indesejáveis (pela definição de classes de comportamentos indesejados com

propriedades e relacionamentos que descrevem características e padrões associados a estes comportamentos);

5. desenvolvimento de sistemas de alerta precoce: que notificam as autoridades competentes sobre possíveis crimes online contra crianças e adolescentes, pode usufruir das vantagens das ontologias uma vez que essas podem definir classes e regras de eventos de alerta, análise de dados de registros, monitoramento de compartilhamento de dados e outros.

Obviamente que estes são apenas alguns exemplos da aplicação de ontologias em diversos cenários de interesse, mas seu uso pode ter um espectro mais amplo graças ao conhecimento de domínio armazenado em suas relações.

7. Considerações finais

Visando facilitar a busca de termos e palavras relacionados ao domínio da violência sexual online contra crianças e adolescentes, neste artigo foi apresentado uma ontologia especialmente concebida para reunir os termos e as relações semânticas entre estes termos e expressões, assim como a metodologia utilizada para seu desenvolvimento. Ressalta-se que a abordagem envolvendo a sequência dicionário semântico, ontologia, e bancos de dados aproxima profissionais e pesquisadores de área distintas para contribuir com a construção e enriquecimento ontológico de termos e definições.

A ontologia aqui proposta – até onde se sabe, inédita – surge como primeiro compêndio mundial que reúne toda a terminologia relativa a crimes sexuais online contra crianças e adolescentes. A instituição Internet Hotline Providers in Europe (INHOPE) que é uma rede internacional para denúncia de conteúdo de abuso sexual infantil na Internet e inclui membros de diversos países divulgou em março de 2023 o lançamento de um projeto de construção de uma ontologia específica para classificação de imagens que “formará a base de um sistema para determinar a ilegalidade de uma imagem em uma determinada jurisdição, independentemente do esquema de categorização e da legislação” [24]. Ao integrar o vocabulário com relações semânticas na forma de um documento digital compatível com padrões da Web Semântica, este constitui-se em uma formalização do conhecimento da área em formato processável.

A ontologia é composta por 925 classes, das quais 696 são subclasses, com 8.371 axiomas, dos quais 3.679 descrevem os relacionamentos entre conceitos. Das 925 classes, 437 foram declaradas como “sameAs”, ou seja, como sinônimas, incluindo em alguns casos versões alternativas em língua portuguesa e em outros a relação com termos em inglês considerados altamente frequentes dentro do contexto do tema.

Por sua abrangência, esta ontologia pode auxiliar no processo de organização e análise de informações relacionadas ao seu domínio, permitindo a identificação de padrões, tendências e áreas de risco, bem como a criação de modelos preditivos e sistemas de recomendação para prevenção e suporte a vítimas.

Considerando-se as análises de todas as atividades pautadas e desenvolvidas, descritas neste artigo, destacam-se que um dos grandes impactos desta pesquisa deverá ser a mobilização para que iniciativas nacionais sejam realizadas em diversos centros, principalmente, em locais onde estudos sejam publicados internacionalmente sobre violência sexual online contra crianças e adolescentes, com base semântica em uma ontologia, uma vez que não foram identificados estudos científicos sobre o assunto nas bases de periódicos. As diversas cartilhas, manuais, apostilas e livros – amplamente produzidos no Brasil – estão distribuídos de forma desordenada por diversos Ministérios Públicos, a exemplo do Ceará, Paraná e Rio de Janeiro, ou repositórios institucionais como UFPR e UFRJ, ou instituições com foco específico na problemática. Com a disponibilização do dicionário semântico e da ontologia, espera-se estimular a pesquisa científica na área de proteção e combate à violência sexual contra crianças e adolescentes para a concepção de políticas públicas pautadas nos conhecimentos produzidos no Brasil e no mundo.

Como trabalho futuro, além da validação e avaliação da Child-Safe, está prevista a adaptação dessa para o uso de alguma ontologia de topo que descreva terminologia e relacionamentos a

respeito de violência numa concepção mais ampla e abstrata. Esta nova ontologia de topo poderia, então, fundamentar esta e diversas outras categorias de violências, potencializando novos esforços na sua compreensão, combate e prevenção.

Finalmente, dado o grande número de classes existentes, geradas a partir das convenções do Produtor (aplicativo escrito em Python), pretende-se reavaliar a estratégia de declaração de classes e seus termos alternativos. Na mesma oportunidade será reavaliada a convenção que declara subclasses e instâncias.

Agradecimentos

Agradecimentos ao Programa das Nações Unidas para o Desenvolvimento (PNUD), e a Universidade Federal do Paraná - UFPR, com a interveniência da Fundação de Apoio à Universidade Federal do Paraná - Fundação de Pesquisas Florestais (FUPEF) e Ministério da Mulher, da Família e dos Direitos Humanos (atual Ministério dos Direitos Humanos e Cidadania) nos termos do Projeto de Execução Nacional BRA/10/007 "Boas práticas na implantação e implementação dos sistemas de informação para a infância e a adolescência - SIPIA, Conselhos tutelares e SIPIA, SINASE WEB".

Referências

- [1] Brasil. Estatuto da Criança e do Adolescente. Lei 8.069, de 13 de julho de 1990, Brasil, http://www.planalto.gov.br/ccivil_03/leis/l8069.htm%0A (1990).
- [2] Brasil. Lei Geral de Proteção de Dados Pessoais. Lei 13.709, de 14 de agosto de 2018, Brasil, http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm%0A (2018).
- [3] Breitman KK. *Web semântica - a internet do futuro*. Rio de Janeiro: LTC, 2005.
- [4] Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis* 1993; 5: 199–220.
- [5] Studer R, Benjamins VR, Fensel D. Knowledge engineering: principles and methods. *Data Knowl Eng* 1998; 25: 161–197.
- [6] Maedche A, Motik B, Stojanovic L, et al. Ontologies for enterprise knowledge management. *IEEE Intell Syst* 2003; 18: 26–33.
- [7] Noy NF, McGuinness DL. *Ontology development 101: a guide to creating your first ontology*, <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html> (2001).
- [8] Khan L, Luo F. Ontology construction for information selection. In: *14th IEEE International Conference on Tools with Artificial Intelligence*. 2002, pp. 122–127.
- [9] Brasil. Constituição da República Federativa do Brasil. Brasil, http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm%0A (1988).
- [10] Brasil. Código Penal. Decreto-lei nº 2.848, de 7 de dezembro de 1940, Brasil, http://www.planalto.gov.br/ccivil_03/Decreto-Lei/Del2848.htm%0A (1940).
- [11] Comitê Gestor da Internet, <https://www.cgi.br/sobre/%0A> (accessed 19 May 2023).
- [12] Safernet, <https://new.safernet.org.br> (accessed 19 May 2023).
- [13] Cetic.br. *Tic Kids Brasil*, <https://cetic.br/pt/pesquisa/kids-online/indicadores/> (accessed 19 May 2023).
- [14] Brasil. Código Internacional de Doenças - 10, <https://www.medicinanet.com.br/cid10.htm> (accessed 19 May 2023).
- [15] INTERPOL. Terminology guidelines for the protection of children from sexual exploitation and sexual abuse, <https://www.interpol.int/Crimes/Crimes-against-children/Appropriate-terminology> (2016, accessed 19 May 2023).
- [16] Heylen K, De Hertog D. Handbook of Terminology Online. John Benjamins Publishing Company, pp. 203–221.
- [17] Pollock D, Peters MD, Khalil H, et al. Recommendations for the extraction, analysis, and

- presentation of results in scoping reviews. *JBI Evid Synth* 2023; 3: 520–532.
- [18] Bibliometrix. Biblioshiny, <https://www.bibliometrix.org/home/index.php/layout/biblioshiny> (2023, accessed 17 July 2023).
- [19] Almeida JPA, Guizzardi G, Sales TP, et al. gUFO: A Lightweight Implementation of the Unified Foundational Ontology (UFO), <http://purl.org/nemo/doc/gufo> (2019).
- [20] Bechhofer S, van Harmelen F, Hendler J, et al. *OWL Web Ontology Language Reference*, <https://www.w3.org/TR/owl-ref/> (2004).
- [21] Webber J. A programmatic introduction to neo4j. *Proc 3rd Annu Conf Syst Program Appl Softw Humanit* 2012; 217–218.
- [22] NEO4J. NEO4J Neosemantics, <https://neo4j.com/labs/neosemantics/4.0/> (2021, accessed 17 July 2023).
- [23] Neto AM. *Ontologias na representação do conhecimento: uma ferramenta semântica para a ciência da informação*. Universidade Estadual Paulista, https://repositorio.unesp.br/bitstream/handle/11449/155929/marinneto_a_me_mar.pdf?sequence=3&isAllowed=y (2018).
- [24] INHOPE. The Global Standard project, <https://www.inhope.org/EN/articles/the-global-standard-project?locale=en> (2023, accessed 17 July 2023).
- [25] UNICEF. Cyberbullying: o que é e como pará-lo, <https://www.unicef.org/brazil/cyberbullying-o-que-eh-e-como-para-lo#resposta9> (accessed 17 April 2023).