

Uma Ontologia Baseada em ML-Schema para Proveniência e Governança de Fluxos de Machine Learning em Educação Superior

Abílio Nogueira Barros¹, Andreza Alencar²,
Fred Freitas³, Alexandre Magno Andrade Maciel¹

¹ Universidade de Pernambuco (UPE)

² Universidade Federal Rural de Pernambuco (UFRPE)

³ Centro de Informática – Universidade Federal de Pernambuco (UFPE)

abilionbarros@gmail.com, andreza.leite@ufrpe.br,
fred@cin.ufpe.br, alexandre.maciel@upe.br

Resumo. A heterogeneidade de abordagens em estudos de machine learning no ensino superior evidencia a necessidade de mecanismos de padronização e rastreabilidade semântica sobre fluxos analíticos educacionais. Este trabalho propõe a *OntoMLEdu*, uma ontologia de aplicação que estende o vocabulário ML-Schema para o domínio do ensino superior, formalizando em Web Ontology Language (OWL 2) conceitos de proveniência, governança de dados e rastreabilidade de Machine Learning Operations (MLOps). Uma prova de conceito demonstrou a capacidade da ontologia de representar ciclos completos de mineração de dados educacionais, identificar lacunas de conformidade de forma autônoma e atuar como camada semântica para trabalhos futuros.

Abstract. The heterogeneity of machine learning approaches in higher education highlights the need for standardization and semantic traceability mechanisms over educational analytical workflows. This paper proposes *OntoMLEdu*, an application ontology that extends the ML-Schema vocabulary to the higher education domain, formally representing in Web Ontology Language (OWL 2) the concepts of data provenance, data governance, and Machine Learning Operations (MLOps) traceability. A proof of concept demonstrated the ontology's ability to represent complete educational data mining workflows, autonomously identify compliance gaps, and serve as a semantic layer for future work.

1. Introdução

Atualmente, observa-se uma crescente disponibilidade de dados em diferentes áreas do setor público, como educação, saúde e mobilidade urbana. Nesse contexto, a tomada de decisão por parte dos gestores demanda não apenas o acesso a esses dados, mas também sua organização de forma estruturada e coerente, bem como a utilização de modelos analíticos e preditivos capazes de identificar tendências e subsidiar a manutenção e o aprimoramento dos serviços ofertados [Ferreira et al. 2021].

No contexto educacional, tanto na educação básica quanto no ensino superior, estudos têm sido desenvolvidos com o objetivo de analisar e prever fenômenos relacionados à trajetória acadêmica dos estudantes, como evasão e retenção. Em especial, a

evasão em cursos de graduação tem sido apontada como um dos principais desafios para a gestão acadêmica das Instituições de Ensino Superior, motivando a aplicação de diferentes técnicas de aprendizagem de máquina para sua predição [Kantorski et al. 2023]. Entretanto, a diversidade de abordagens adotadas e a heterogeneidade na construção desses modelos evidenciam a necessidade de mecanismos que favoreçam sua descrição, padronização e reutilização. Nesse sentido, torna-se relevante a proposição de uma estrutura capaz de organizar, descrever e integrar tais modelos, promovendo maior interoperabilidade entre soluções e ampliando seu potencial de reaproveitamento em diferentes contextos educacionais.

O estabelecimento de mecanismos que promovam a organização semântica tanto dos dados quanto dos modelos analíticos mostra-se fundamental para viabilizar sua integração e reutilização em diferentes contextos. Ontologias, enquanto instrumentos formais de representação do conhecimento, apresentam-se como uma alternativa promissora para estruturar conceitos, relações e restrições de um domínio específico, permitindo a construção de camadas semânticas capazes de mediar a comunicação entre diferentes sistemas e aplicações. No domínio educacional, ontologias têm sido propostas para representar formalmente estruturas de conhecimento pedagógico, como a classificação de objetivos educacionais [de Oliveira et al. 2020], como também para o mapeamento e integração de avaliações do ensino superior na UFRGS [Daniel et al. 2023].

Diante desse cenário, este trabalho propõe uma ontologia para o contexto do ensino superior, destinada à padronização da representação de dados educacionais e à estruturação da interação com modelos de *machine learning*. A ontologia atua como uma camada semântica de interoperabilidade, capaz de descrever entidades, indicadores, atributos e fluxos de dados associados a processos analíticos, favorecendo a integração, rastreabilidade e reutilização de recursos em ambientes de análise educacional.

2. Contextualização e Trabalhos relacionados

Uma ontologia pode ser compreendida de diversas formas e sob diferentes perspectivas, seja no campo filosófico, literário ou puramente semântico. No contexto da representação semântica, uma ontologia tem por finalidade definir e representar o mini-mundo ao qual se propõe a mapear, estruturando esse domínio por meio de relações, restrições e axiomas formais [Sure et al. 2009].

As ontologias podem ser classificadas de acordo com seu nível de generalidade e escopo de aplicação. As ontologias de topo, ou fundacionais, descrevem conceitos abstratos e independentes de domínio, como evento, objeto, tempo e espaço, servindo como base para ontologias mais especializadas [Uschold and Gruninger 1996]. Em um nível intermediário, as ontologias core capturam conceitos compartilhados por conjuntos de domínios relacionados, sem se restringirem a uma área específica [Oberle et al. 2005].

Já as ontologias de domínio modelam conceitos próprios de campos particulares do conhecimento, como medicina, direito ou engenharia, enquanto as ontologias de aplicação descrevem conceitos voltados a tarefas ou sistemas específicos, geralmente estendendo ontologias de domínio para atender requisitos operacionais. Nesse contexto, a reutilização de ontologias constitui um princípio fundamental da modelagem semântica, permitindo o aproveitamento de conceitos e entidades previamente formalizados por diferentes comunidades, reduzindo o esforço de desenvolvimento, promovendo a interope-

tabilidade semântica e favorecendo maior consistência entre representações de domínios correlatos [Noy et al. 2001].

A utilização de ontologias no contexto educacional encontra-se consolidada na literatura, abrangendo diferentes aplicações relacionadas ao ensino e à gestão acadêmica. No âmbito do ensino, [Passos et al. 2023] desenvolveram uma ontologia voltada à aprendizagem de robótica educacional utilizando a metodologia *Ontology Development 101*, com apoio de especialistas na elaboração de *competency questions* e validação por meio de consultas *SPARQL*. De forma complementar, [Daniel et al. 2025] propuseram uma ontologia para modelar sistemas de avaliação no ensino superior, integrando métricas institucionais e técnicas de Processamento de Linguagem Natural (PLN) para extração de sentimentos a partir de dados qualitativos, possibilitando a representação conjunta de informações quantitativas e subjetivas em uma rede semântica.

No contexto da predição de evasão acadêmica, [da Silva et al. 2022] apresentaram uma ontologia na qual a modelagem semântica constitui a etapa inicial de um processo mais amplo de análise educacional. Nesse modelo, técnicas de *machine learning* são empregadas para realizar predições relacionadas à evasão discente, cujos resultados são posteriormente disponibilizados por meio de uma *Application Programming Interface* (API), demonstrando o potencial das ontologias como mecanismo de integração entre dados educacionais e sistemas analíticos.

O diferencial deste trabalho em relação às abordagens anteriores reside na proposição de uma ontologia que, além de representar conceitos do domínio educacional, modela de forma integrada os dados de entrada e o ciclo de vida dos modelos analíticos, desde sua configuração e treinamento até a disponibilização dos resultados. Essa ontologia será uma extensão da ontologia de domínio ML-schema.

3. Metodologia

A metodologia adotada para o desenvolvimento da ontologia proposta neste trabalho é diretamente derivada do guia *Ontology Development 101*, publicado pela Stanford University [Noy et al. 2001]. Tal guia propõe uma abordagem iterativa e incremental para a criação de ontologias, na qual o processo deve refletir, da forma mais fidedigna possível, o mini-mundo que a ontologia se propõe a descrever.

O desenvolvimento é conduzido em ciclos sucessivos de refinamento, permitindo que o modelo semântico evolua de maneira controlada e coerente com os requisitos do domínio. O processo prático de desenvolvimento envolve quatro etapas centrais, executadas de forma cíclica e progressivamente refinadas ao longo do projeto, conforme os seguintes passos: (i)Definição,(ii)Integração,(iii)Modelagem e (iv)Instanciação.

3.1. Definição

Essa estabeleceu o escopo e a finalidade central da ontologia, delimitando seu propósito como um artefato de auditoria e rastreabilidade focado na implementação de um *Escudo Sanitário*. Nesta fase, foram formuladas as Perguntas de Competência (*Competency Questions*) (CQs), responsáveis por guiar toda a modelagem e assegurar que o grafo de conhecimento fosse capaz de mapear o ciclo de vida da informação desde a extração nos sistemas acadêmicos transacionais até a exportação do modelo preditivo, garantindo a padronização e o isolamento dos dados.

As Perguntas de Competência atuam como um contrato de requisitos para a ontologia: delimitam seu escopo e determinam que, caso o grafo de conhecimento não seja capaz de respondê-las por meio de uma linguagem de consulta formal (como SPARQL), o modelo está incompleto [Noy et al. 2001]. Para a ontologia proposta, denominada OntoMLEdu, as CQs foram agrupadas em três pilares arquiteturais, conforme descrito a seguir:

Tabela 1. Perguntas de Competência definidas para a OntoMLEdu

Pilar	Pergunta de Competência
Proveniência	Quais fontes de dados originais compõem a base de dados consolidada (<i>StagingDataset</i>) utilizada como entrada para um modelo específico?
Proveniência	Que processos foram aplicados para o lote de dados brutos (<i>RawDataset</i>) para a derivação das bases de dados de treinamento (<i>StagingDataset</i>)?
Governança e Preparo dos Dados	A qual processo de anonimização a base de dados submetida, o <i>pipeline</i> de treinamento passou obrigatoriamente por um processo formal de anonimização (<i>DataAnonymization</i>)?
Governança e Preparo dos Dados	Quais técnicas estruturais de preparação de dados (<i>MissingValueHandling</i> , <i>ClassBalancing</i> e <i>DataNormalization</i>) foram declaradas na construção de um <i>StagingDataset</i> para utilização nos modelos de ML?
Rastreabilidade e Performance	Qual artefato executável (<i>TrainingPipeline</i>) foi utilizado durante o evento de execução (<i>ModelExecution</i>) que produziu um modelo exportado (<i>OutputModel</i>) específico?
Rastreabilidade e Performance	Qual algoritmo (<i>PredictAlgorithm</i>) atingiu o melhor desempenho e quais métricas de validação foram obtidas para uma determinada tarefa educacional?

3.2. Integração

Dedicou-se à avaliação e ao reúso de ontologias preexistentes, princípio estrutural da Web Semântica adotado para assegurar a interoperabilidade e mitigar a sobrecarga arquitetural. A seleção do vocabulário ML-Schema (*mls*) como base fundacional foi metodologicamente justificada pela sua aderência direta às Perguntas de Competência definidas na fase anterior. Uma vez que as CQs exigiam a rastreabilidade minuciosa de todo o ecossistema de MLOps requerendo respostas formais sobre instâncias de execução, *pipelines* de treinamento, algoritmos preditivos e métricas de desempenho, o ML-Schema proveu as superclasses ideais já consolidadas pela comunidade W3C [Publio et al. 2018]. A adoção dessa ontologia base evitou a redundância na modelagem de conceitos universais de inteligência artificial, permitindo que o esforço semântico da OntoMLEdu se concentrasse na extensão desse núcleo para atender aos requisitos específicos do domínio educacional.

3.3. Modelagem

Essa etapa concentrou-se na formalização do dicionário de dados e na organização da hierarquia taxonômica. Os conceitos foram categorizados alinhando as entidades do domínio

educacional e as camadas de engenharia de dados aos termos do ML-Schema por meio de relações de herança semântica. Esta etapa também assegurou a integridade estrutural do modelo ao estabelecer regras de exclusão mútua e disjunção de classes, proibindo, por exemplo, a sobreposição conceitual entre o estado imutável das instâncias de extração bruta (Raw) e as bases de modelagem processadas e compostas por *features* (Staging).

Para além do mapeamento das estruturas de dados, a modelagem expandiu-se para englobar todo o ciclo de vida analítico, estendendo os termos de topo do ML-Schema por meio de relações de herança semântica. Dessa forma, formalizaram-se subclasses específicas para representar tarefas educacionais, *pipelines* de treinamento, eventos de execução, modelos matemáticos exportados e métricas globais de avaliação.

3.4. Instanciação

A Etapa 4 concentrou-se na formalização do comportamento lógico e na validação do modelo semântico por meio da definição de *Object Properties*, *Data Properties*, axiomas de restrição e da instanciação da ontologia com *Individuals*. Nessa fase, foram estabelecidos os relacionamentos entre as classes, seus domínios e contradomínios, incluindo mecanismos de rastreabilidade, como a ligação entre conjuntos de dados tratados e suas respectivas fontes brutas. Adicionalmente, foram implementadas restrições em OWL que compõem o *Escudo Sanitário* do *framework*, impondo requisitos de governança, como a obrigatoriedade de processos de anonimização para bases destinadas ao treinamento. Por fim, a instanciação de cenários extraídos da literatura permitiu validar a capacidade da ontologia de auditar e inferir automaticamente a conformidade de fluxos de Mineração de Dados Educacionais (MDE).

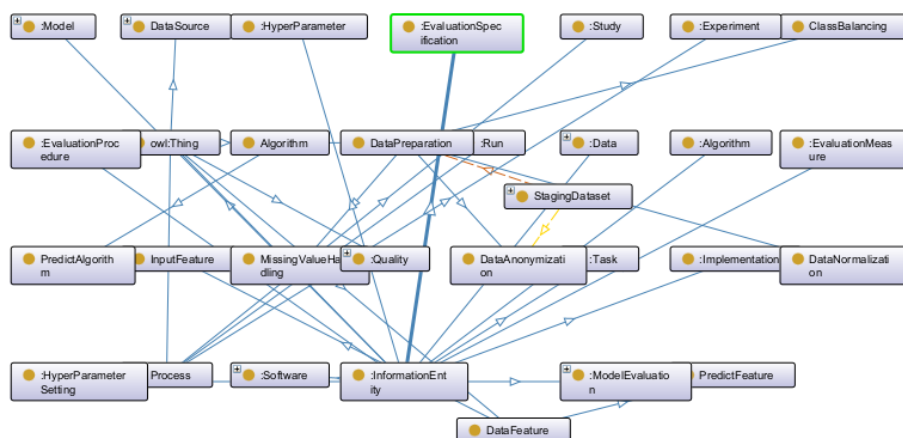


Figura 1. Grafo das entidades da OntoMLEdu.

4. OntoMLEdu: A formalização da ontologia

A ontologia OntoMLEdu (*Ontology for Machine Learning in Education*) foi concebida como uma ontologia de aplicação que estende formalmente o vocabulário ML-Schema (mls), assegurando interoperabilidade com o padrão semântico consolidado pela comunidade W3C para a representação de processos de aprendizado de máquina. Sua construção lógica, orientada por uma adaptação da metodologia *Ontology Development 101* ao contexto de MDE e governança de MLOps, concentrou-se na especialização dos conceitos genéricos de aprendizado de máquina para atender requisitos de auditoria, proveniência e rastreabilidade no domínio educacional.

4.1. Hierarquia de Classes

A taxonomia da OntoMLEdu é organizada em três eixos semânticos principais: **gestão de dados**, **ciclo de vida do modelo** e **tarefas educacionais**. Cada eixo é composto por classes próprias da ontologia ou por extensões diretas de superclasses do ML-Schema, conforme a sequência a seguir:

O eixo de gestão de dados é estruturado pelas classes `RawDataset` e `StagingDataset`, ambas subclasses de `mls`. A classe `RawDataset` representa dados extraídos diretamente das fontes institucionais e é declarada como disjunta de `StagingDataset` (owl), formalizando o *Escudo Sanitário* ao impedir que um mesmo indivíduo pertença simultaneamente às zonas bruta e tratada. Adicionalmente, `RawDataset` possui uma restrição de cardinalidade qualificada que exige exatamente uma relação `hasOriginalSource` com `DataSource`, garantindo a rastreabilidade da proveniência dos dados.

Por sua vez, `StagingDataset` possui duas restrições existenciais: ao menos uma relação `isDerivedFromRaw` com um `RawDataset`, assegurando a linhagem dos dados processados, e ao menos uma relação `hasPreparationStep` com `DataAnonymization`, tornando a anonimização uma pré-condição lógica para qualquer conjunto de dados destinado ao treinamento. A classe `DataSource` é especializada nas subclasses disjuntas `PublicDataSource` e `PrivateDataSource`, representando, respectivamente, repositórios abertos e sistemas institucionais restritos. Complementarmente, `DataPreparation` atua como superclasse das etapas de pré-processamento, sendo especializada em `DataAnonymization`, `ClassBalancing`, `MissingValueHandling` e `DataNormalization`, contemplando os principais processos de engenharia de dados do domínio educacional.

Já o ciclo de vida do modelo é representado por três classes que estendem o ML-Schema. A classe `TrainingPipeline`, subclasse de `mls:Software`, representa o artefato executável responsável pela condução do treinamento. A classe `ModelExecution`, subclasse de `mls:Model`, registra o evento de execução e possui um axioma de restrição universal (owl:allValuesFrom) sobre a propriedade `mls:hasInput`, determinando que toda execução de modelo deve consumir exclusivamente instâncias de `StagingDataset` vedando formalmente o acesso direto a dados da zona bruta e consolidando, em OWL, a principal regra arquitetural do *Escudo Sanitário*. A classe `OutputModel`, também subclasse de `mls:Model`, representa o artefato preditivo exportado ao final do processo de treinamento.

A rastreabilidade entre o evento de execução e seu artefato executável é estabelecida pela *Object Property* `usesExecutable`, cujo domínio é `mls:Model` e cujo contradomínio é `TrainingPipeline`. Seguindo, temos a classe `PredictModel`, subclasse de `mls:Run`, representa a execução concreta de um modelo preditivo, enquanto `GlobalMetrics`, subclasse de `mls:ModelEvaluation`, formaliza o registro das métricas globais de avaliação como F1-Score e Precision, viabilizando consultas comparativas de desempenho entre diferentes configurações de um mesmo modelo ou diferentes versões de modelos para uma mesma tarefa.

Por fim, o eixo de Tarefas Educacionais e Algoritmos é estruturado a partir da classe `EducationalTask`, subclasse de `mls`, que especializa o conceito de tarefa para

o domínio educacional, permitindo representar formalmente objetivos analíticos como predição de evasão, classificação de risco acadêmico e análise de retenção. A modelagem dos algoritmos é realizada por meio da classe *Algorithm* e de sua especialização *PredictAlgorithm*, destinada aos algoritmos preditivos empregados nessas tarefas. Complementarmente, a representação das variáveis do modelo é realizada pela classe *DataFeature*, especializada em *InputFeature*, correspondente às variáveis de entrada, e *PredictFeature*, correspondente às variáveis-alvo ou de saída das predições.

4.2. Propriedades de Objeto

A *OntoMLEdu* define quatro *Object Properties* originais, complementares às propriedades herdadas do *ML-Schema*, conforme sintetizado na Tabela 2.

Tabela 2. Object Properties definidas na *OntoMLEdu*

Propriedade	Domínio	Contradomínio
<code>hasOriginalSource</code>	<code>RawDataset</code>	<code>DataSource</code>
<code>isDerivedFromRaw</code>	<code>StagingDataset</code>	<code>RawDataset</code>
<code>hasPreparationStep</code>	<code>StagingDataset</code>	<code>DataPreparation</code>
<code>usesExecutable</code>	<code>mls:Model</code>	<code>TrainingPipeline</code>

A propriedade `hasStorageURI`, associada ao domínio `mls:Data`, é declarada com restrição de tipo `xsd:anyURI`, permitindo o armazenamento formal do endereço físico de cada artefato de dado no *triplestore*.

4.3. Axiomas de Restrição e Coerência Lógica

Os axiomas construídos constituem o mecanismo central de validação lógica da ontologia. Em síntese, três regras estruturais garantem a coerência do modelo: (i) a disjunção entre `RawDataset` e `StagingDataset`, que impede a sobreposição entre zonas de dados; (ii) a restrição existencial de `DataAnonymization` como pré-condição obrigatória de `StagingDataset`, que vincula qualquer base de treinamento válida a um processo formal de anonimização; e (iii) a restrição universal sobre `mls:hasInput` em `ModelExecution`, que proíbe o consumo direto de dados brutos por qualquer evento de execução de modelo. Essas três restrições, quando ativadas por um *reasoner* OWL como o *HermiT*, permitem a inferência automática de violações arquiteturais, conferindo à ontologia capacidade autônoma de auditoria sobre o fluxo de MDE.

4.4. Prova de Conceito

Para avaliar a capacidade representacional da ontologia *OntoMLEdu*, foi conduzida uma prova de conceito baseada no trabalho de [Kantorski et al. 2023], intitulado *MDE para Predição da Evasão em Cursos de Graduação Presenciais no Ensino Superior*, publicado nos Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE 2023). O trabalho foi selecionado por apresentar um fluxo metodológico completo e explicitamente descrito, desde a extração dos dados brutos institucionais até a avaliação comparativa de múltiplos modelos preditivos, o que o torna um caso de uso ideal para a validação teórica da ontologia criada.

A aplicação da *OntoMLEdu* ao trabalho de [Kantorski et al. 2023] resultou em um mapeamento que cobre integralmente o fluxo de MDE descrito pelos autores, desde

a extração dos dados institucionais até a exportação e avaliação dos modelos preditivos. Das seis Perguntas de Competência definidas na Etapa 1 da metodologia, cinco foram respondidas com plena conformidade, uma foi classificada como parcialmente conforme e uma evidenciou uma lacuna de governança, conforme sintetizado na Tabela 3.

Tabela 3. Conformidade das Perguntas de Competência — Mapeamento de Kantorski et al. (2023) à OntoMLEdu

CQ	Classe / Propriedade OntoMLEdu	Evidência em Kantorski et al. (2023)	Conformidade
CQ1	PrivateDataSource → hasOriginalSource → RawDataset	Dados extraídos do sistema acadêmico da UFSM (106 cursos · 3 semestres)	Conforme
CQ2	StagingDataset → isDerivedFromRaw → RawDataset	Conjunto bruto transformado com seleção de atributos por correlação (CFS)	Conforme
CQ3	owl:someValuesFrom DataAnonymization em StagingDataset	Etapa de anonimização ausente no trabalho; <i>reasoner</i> sinalizaria inconsistência	Lacuna
CQ4	MissingValueHandling via hasPreparationStep	Tratamento de valores ausentes descrito; balanceamento e normalização não declarados	Parcial
CQ5	ModelExecution → usesExecutable → TrainingPipeline	Pipeline baseado em CRISP-DM vinculado a cada execução por algoritmo	Conforme
CQ6	PredictAlgorithm → ModelExecution → GlobalMetrics	Acurácias de 85% a 96% por algoritmo; instâncias consultáveis via SPARQL	Conforme

5. Conclusão e trabalhos futuros

O presente trabalho apresentou a OntoMLEdu, uma ontologia de aplicação baseada no ML-Schema e voltada à interoperabilidade entre dados educacionais e modelos de *machine learning*, incorporando conceitos de proveniência, governança de dados, rastreabilidade de MLOps e avaliação de desempenho preditivo. A prova de conceito realizada sobre o estudo de [Kantorski et al. 2023] demonstrou a capacidade da ontologia de representar formalmente o ciclo de vida de processos de MDE, responder à maior parte das Perguntas de Competência propostas e identificar automaticamente uma lacuna de conformidade relacionada à ausência de anonimização dos dados, evidenciando seu potencial como instrumento de auditoria semântica e apoio à governança.

Como continuidade natural deste trabalho, pretende-se avançar a OntoMLEdu para além de sua função primária de auditoria e rastreabilidade semântica. Em etapas futuras, a ontologia passará a atuar como base de conhecimento estruturada para a implementação de Modelos de Linguagem Reduzidos (*Small Language Models* (SLMs)), nos quais o grafo de conhecimento servirá como repositório formal de contexto, orientando os agentes na geração de respostas fundamentadas nos dados e modelos institucionais previamente validados a camada de conhecimento sobre a qual os agentes de ML apoiarão a tomada de decisão de gestores institucionais, oferecendo recomendações rastreáveis, auditáveis e semanticamente consistentes com o domínio do ensino superior.

6. Declaração de uso de IA

Ferramentas de Inteligência Artificial Generativa foram utilizadas como ferramentas de apoio na elaboração desse trabalho, como no auxílio na revisão textual pela ferramenta Gemini, e criação das tabelas e artefatos ilustrativos apresentadas durante o trabalho com ferramenta Claude.

Referências

- da Silva, E. M., Mutz, F. W., and Ruy, F. B. (2022). Uso de ontologias no suporte a aplicação de machine learning: Um caso no domínio de evasão escolar. In *ONTOBRAS*, pages 32–45.
- Daniel, E. M. P., de Oliveira, J. P., Wives, L. K., and Lorandi, A. (2023). Transformando o processo de avaliação da qualidade no ensino superior: Uma proposta ontológica para a ufrgs. In *Workshop em Estratégias Transformadoras e Inovação na Educação (WETIE)*, pages 62–72. SBC.
- Daniel, E. M. P., Wives, L. K., Lorandi, A., and de Oliveira, J. P. M. (2025). Ontologia para análise qualitativa automatizada em avaliações de cursos de ensino superior. In *Simpósio Brasileiro de Banco de Dados (SBBDD)*, pages 371–381. SBC.
- de Oliveira, P. R. F., de Alencar Silva, P., and Neto, P. F. R. (2020). Ontologia dos objetivos educacionais. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1183–1192. SBC.
- Ferreira, L. A., Rodrigues, R. L., and de Souza, R. N. (2021). Dados abertos educacionais brasileiros: Um mapeamento sistemático da literatura. *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1186–1195.
- Kantorski, G. Z., Martins, R. Z., Balejo, A., and Frick, M. (2023). Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1133–1142. SBC.
- Noy, N. F., McGuinness, D. L., et al. (2001). *Ontology development 101: A guide to creating your first ontology*.
- Oberle, D., Staab, S., and Volz, R. (2005). Three dimensions of knowledge representation in wonderweb. *Künstliche Intell.*, 19(1):31.
- Passos, D. D. N., de Araújo, S. R. F., Silva, S. D., and Queiroz, P. G. G. (2023). Uma ontologia para apoiar o ensino de matemática básica com uso de robótica educacional. *HOLOS*, 8(39).
- Publio, G. C., Esteves, D., Panov, P., Soldatova, L., Soru, T., Vanschoren, J., Zafar, H., et al. (2018). *ML-schema: exposing the semantics of machine learning with schemas and ontologies*. *arXiv preprint arXiv:1807.05351*.
- Sure, Y., Staab, S., and Studer, R. (2009). Ontology engineering methodology. In *Handbook on ontologies*, pages 135–152. Springer.
- Uschold, M. and Gruninger, M. (1996). *Ontologies: Principles, methods and applications*. *The knowledge engineering review*, 11(2):93–136.