

# Ontologia e Microdados do ENADE: Uma Arquitetura Semântica para Identificação de Lacunas de Aprendizagem e Apoio à Recomendação Educacional

## ENADE Ontology and Microdata: A Semantic Architecture for Identifying Learning Gaps and Supporting Educational Recommendations

Ivanylson Honorio Gonçalves<sup>1</sup>, Regina Braga<sup>2</sup>, José Maria David<sup>3</sup>, Victor Stroele<sup>4</sup>

**Resumo:** Este trabalho propõe uma arquitetura para utilizar os microdados do ENADE em um sistema de recomendação de conteúdo educacional. A solução se baseia na criação de uma ontologia que organiza e categoriza semanticamente as informações sobre cursos, competências e objetos de conhecimento. A prova de conceito detalhada neste artigo valida a construção da ontologia e demonstra sua capacidade de inferir lacunas de conhecimento a partir dos dados de desempenho brutos, estabelecendo o alicerce semântico para a personalização. Este estudo estabelece o alicerce semântico fundamental que ultrapassa as limitações da análise estatística convencional, sendo um passo essencial para a criação de futuras plataformas de apoio pedagógico e recomendação educacional inteligente integrando técnicas de machine learning e modelos colaborativos.

**Palavras-chave:** ENADE. Ontologia. Sistema de Recomendação. Microdados Educacionais. Objeto de Conhecimento. Análise de Dados Educacionais. Web Semântica. Machine Learning. Processamento Inteligente de Dados. Educação Superior no Brasil.

**Abstract:** This work proposes an architecture for using ENADE microdata in an educational content recommendation system. The solution is based on the creation of an ontology that semantically organizes and categorizes information about courses, competencies, and knowledge objects. The proof of concept detailed in this article validates the ontology's construction and demonstrates its ability to infer knowledge gaps from raw performance data, establishing the semantic foundation for personalization. This study establishes the fundamental semantic foundation that overcomes the limitations of conventional statistical analysis, representing an essential step towards the creation of future platforms for pedagogical support and intelligent educational recommendation, integrating machine learning techniques and collaborative models.

**Keywords:** ENADE. Ontology. Recommendation System. Educational Microdata. Knowledge Object. Educational Data Analysis. Semantic Web. Machine Learning. Intelligent Data Processing. Higher Education in Brazil.

<sup>1</sup> Mestrando do Programa de Pós-graduação em Ciência da Computação (Universidade Federal de Juiz de Fora), E-mail: ivanylson.honorio@estudante.ufjf.br

<sup>2</sup> Doutora em Engenharia de Sistemas e Computação pela UFRJ, docente e pesquisadora do Departamento de Ciência da Computação e do Programa de Pós-graduação em Ciência da Computação (Universidade Federal de Juiz de Fora), E-mail: regina.braga@ufjf.br.

<sup>3</sup> Doutor em Engenharia de Sistemas e Computação pela UFRJ, docente e pesquisador do Departamento de Ciência da Computação e do Programa de Pós-graduação em Ciência da Computação (Universidade Federal de Juiz de Fora), E-mail: jose.david@ufjf.br.

<sup>4</sup> Doutor em Engenharia de Sistemas e Computação pela UFRJ, docente e pesquisador do Departamento de Ciência da Computação e do Programa de Pós-graduação em Ciência da Computação (Universidade Federal de Juiz de Fora), E-mail: victor.stroele@ufjf.br.

# 1. Introdução

O Exame Nacional de Desempenho dos Estudantes (ENADE) é uma ferramenta essencial para a avaliação da qualidade do ensino superior no Brasil (GOV.BR, 2025). Através da análise dos microdados fornecidos pelo exame, é possível obter dados sobre o desempenho dos estudantes, permitindo que as instituições de ensino, como a Universidade Federal de Juiz de Fora (UFJF) no contexto deste trabalho, compreendam e aprimorem seus processos pedagógicos. A análise de dados educacionais, portanto, é um pilar fundamental para a melhoria contínua do ensino e da aprendizagem.

Apesar da riqueza de informações contida nos microdados do ENADE, sua análise e utilização apresentam desafios significativos (Cruz, 2023). A principal dificuldade reside na forma como os objetos de conhecimento das questões estão estruturados, o que pode dificultar a identificação de lacunas de aprendizado identificadas no exame de maneira precisa e acionável. A transformação desses dados brutos em informações úteis e personalizadas requer uma abordagem mais sofisticada do que a simples análise estatística, o que justifica a necessidade de uma arquitetura inteligente para processamento e interpretação.

Neste sentido, este trabalho tem como objetivo principal deste trabalho propor e validar uma arquitetura semântica que utiliza os microdados do ENADE como insumo. A solução centraliza-se na criação de uma ontologia, que funciona como uma estrutura formal para organizar semanticamente as informações sobre cursos, competências e objetos de conhecimento. Ao transformar os dados brutos e despadronizados em conhecimento estruturado, a ontologia viabiliza a identificação exata das lacunas de aprendizagem através de regras de inferência lógica. Desta forma, o estudo estabelece o alicerce semântico indispensável para a personalização, servindo como a base para futuros sistemas de recomendação de conteúdo educacional.

Este artigo está organizado em cinco seções principais. Na seção 2, Fundamentação Teórica, são abordados conceitos fundamentais para o entendimento do trabalho. A seção 3 discute os trabalhos correlatos; a seção 4 o estado da arte e originalidade deste trabalho; a seção 5 descreve a metodologia utilizada nesta pesquisa; a seção 6 detalha a arquitetura proposta; na seção 7 detalha questões tecnológicas utilizadas e decisões de projeto na implementação da proposta inicial; seção 8 a demonstração e avaliação funcional da ontologia; seção 9 discussão sobre a ontologia e arquitetura. Por fim, nas considerações finais, são sintetizados os principais avanços e contribuições do trabalho que servirá de base para o futuro sistema de recomendação e análise educacional, além de apontar perspectivas para desenvolvimentos futuros.

## 2. Fundamentação Teórica

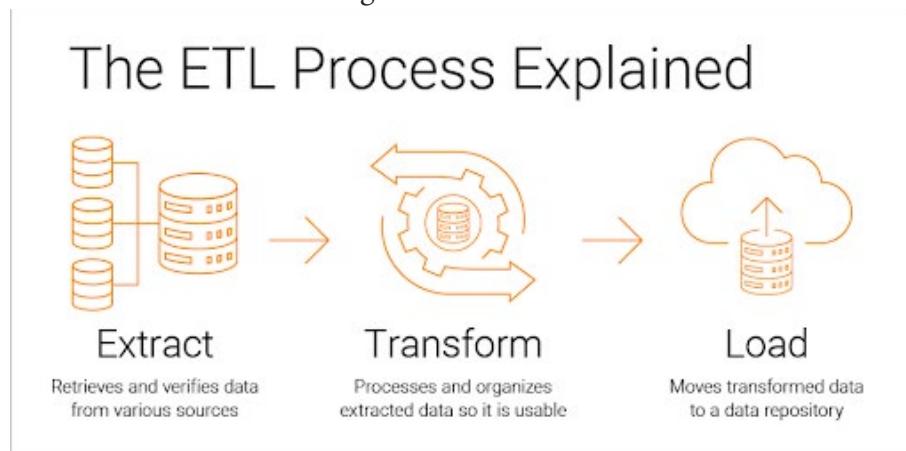
O Exame Nacional de Desempenho dos Estudantes (ENADE) é uma ferramenta essencial para avaliar a qualidade do ensino superior no Brasil. Com base no tripé do Sistema Nacional de Avaliação da Educação Superior (SINAES) (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2025), busca mensurar o desempenho de concluintes em relação às competências e habilidades previstas nos currículos dos cursos superiores, composto por questões objetivas e discursivas que avaliam tanto o conhecimento específico quanto o geral dos estudantes. Além disso, o ENADE desempenha um papel estratégico ao fornecer dados que ajudam a monitorar a evolução da educação superior no país (De Oliveira, 2023). Os microdados gerados pelo ENADE fornecem informações sobre o desempenho em diferentes áreas do conhecimento. Desde sua implementação, o ENADE passou por diversas mudanças metodológicas que aprimoraram sua capacidade de avaliação. Essas alterações refletem o esforço contínuo em capturar de forma mais precisa a qualidade dos cursos e o impacto do processo formativo sobre os estudantes. Atualmente, o exame é uma referência nacional e internacional no contexto de avaliação educacional (Belo, 2017).

Os objetos de conhecimento, conforme descrito pela Base Nacional Comum Curricular (BNCC) (INEP, 2016), são um alicerce fundamental para a análise e categorização das questões do ENADE. A associação entre questões e objetos de conhecimento permite identificar com precisão áreas de melhoria nos currículos e estratégias pedagógicas. Este processo promove uma integração entre análise quantitativa e qualitativa, permitindo ações mais efetivas para o aprimoramento educacional (Choji, 2021). O conceito de dados abertos é central para trabalharmos com os dados

do ENADE, pois possibilita o acesso irrestrito a informações para análises (PORTAL BRASILEIRO DE DADOS ABERTOS, 2025). A transparência e a padronização proporcionadas pelos dados abertos são fundamentais para a criação de soluções inovadoras, como as abordadas neste estudo. O uso de dados abertos educacionais, alinhado às diretrizes legais brasileiras, garante uma base para investigações acadêmicas e para o desenvolvimento de ferramentas práticas de gestão.

Para isso, este trabalho emprega um processo ETL (Extração, Transformação e Carregamento) inspirado no método descrito por Rodrigues e Maciel (2022), que propõem um procedimento para capturar e compartilhar dados abertos educacionais, garantindo acessibilidade tanto para máquinas quanto para humanos. Esse processo é adaptado para atender às especificidades da análise dos microdados do ENADE 2022 no contexto dos cursos da Universidade Federal de Juiz de Fora (UFJF), com o objetivo de identificar lacunas nos objetos de conhecimento e propor recomendações de conteúdo. O processo é dividido em três etapas principais, conforme demonstrado na Figura 1:

Figura 1. Processo ETL



Fonte: <https://www.informatica.com/resources/articles/what-is-etl.html>

- **Extração:** Nesta etapa, os microdados do ENADE são coletados a partir de fontes públicas, como os bancos de dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Esses dados incluem informações detalhadas sobre o desempenho dos estudantes, características dos cursos e contexto educacional, que servirão como base para a análise.
- **Transformação:** Após a extração, os dados brutos são transformados por meio de um mapeamento baseado em modelos de referência e enriquecimento semântico. No presente estudo, essa transformação é essencial para organizar os microdados do ENADE de forma estruturada, categorizando os objetos de conhecimento em categorias ontológicas. O enriquecimento semântico é realizado com base em taxonomias e padrões previamente definidos, o que permite relacionar os dados às lacunas identificadas nos cursos analisados.
- **Carregamento e Publicação:** A última etapa consiste na conversão dos dados estruturados em uma ontologia no formato .owl (Web Ontology Language)<sup>1</sup> (W3C, 2001) (Souza e Alvarenga, 2004). Esse formato é amplamente utilizado para representar informações semânticas, permitindo tanto a consulta automatizada quanto a análise manual. A adaptação do método para este trabalho visa transformar os microdados do ENADE em instâncias de uma ontologia que atenda ao objetivo de categorizar os objetos de conhecimento e propor recomendações baseadas em evidências (Berners-Lee e T. Miller, 2002). Para isso, a ontologia foi construída com a ferramenta Protégé<sup>2</sup>, que permite a criação de classes, propriedades e instâncias representando os conceitos educacionais analisados (Dziekaniak, 2004). O formato .owl é utilizado para garantir a interoperabilidade com sistemas de aprendizado de máquina, facilitando a análise automatizada e a geração de recomendações.

Ontologias são modelos conceituais que estruturam, organizam e compartilham informações em um determinado domínio, promovendo a interoperabilidade entre sistemas (De Lima et al., 2004). No campo educacional,

1 <https://protege.stanford.edu/>

2 <https://www.w3.org/TR/owl-features/>

desempenham um papel crucial ao classificar e relacionar conceitos, facilitando a organização de conteúdos e a identificação de lacunas no aprendizado (Ferreira, 2023). Além disso, integram dados de diferentes fontes, possibilitando buscas mais precisas e inferências lógicas. Sua aplicação é essencial tanto para a Web Semântica quanto para sistemas inteligentes, ao oferecer uma base estruturada para a gestão e análise de conhecimento (Pickler, 2007).

O mapeamento semântico dos dados é um passo crítico para garantir que as informações representadas na ontologia estejam alinhadas com as lacunas educacionais identificadas. Além disso, a ontologia resultante poderá ser iterativamente refinada por meio da curadoria colaborativa, em que docentes e discentes poderão validar e atualizar os conceitos e relações definidos, promovendo um ciclo contínuo de melhoria da qualidade dos dados e das recomendações geradas. Participação de múltiplos atores no processo de seleção, organização e recomendação de conteúdos (Fuks e Pimentel, 2011). Em um ambiente educacional, essa abordagem fomenta a troca de conhecimento entre professores e alunos, promovendo um aprendizado mais dinâmico e efetivo. Esses conceitos fundamentam a proposta apresentada, integrando soluções tecnológicas e pedagógicas para superar os desafios identificados nos microdados do ENADE.

### 3. Trabalhos relacionados

A metodologia adotada nesta pesquisa foi baseada em pesquisa documental e revisão bibliográfica. Para identificar os principais artigos para desenvolver o trabalho, foi necessário realizar pesquisa documental e revisão bibliográfica utilizando diversas fontes, incluindo IEEE Explorer; ACM Digital Library; Citeseer library; SpringerLink; WILEY Interscience; ScienceDirect; e Google Scholar. (para artigos científicos mais citados), livros e sites especializados na temática. As palavras-chave utilizadas foram “ENADE”, “Ontologia ENADE”, “Objeto de conhecimento do ENADE”, “Relatório de Síntese ENADE”. O resultado foi a seleção de alguns trabalhos relevantes relacionados à abordagem proposta.

Diversos pesquisadores e discentes têm realizado estudos envolvendo o uso de microdados do ENADE e relatórios do INEP para analisar instituições de ensino superior, como ilustrado pelos trabalhos de Rendeiro et al. (2023), Charão et al. (2020), Lima et al. (2021), Cunha et al. (2021), Vieira et al. (2022) e Brito (2016). Esses estudos, encontrados por meio da ferramenta Google Acadêmico, destacam-se pelo foco em diferentes abordagens e metodologias para entender e aprimorar o desempenho acadêmico em cursos superiores.

Rendeiro et al. (2023) analisaram o desempenho de estudantes do curso de Engenharia de Computação da Universidade Federal do Pará, utilizando microdados do ENADE de 2014 e 2019. A pesquisa identificou a necessidade de melhorias em temas diretamente relacionados à Ciência da Computação e sugeriu a atualização do processo pedagógico como um todo.

Charão et al. (2020) buscaram subsidiar a reformulação do projeto pedagógico de curso (PPC) do Bacharelado em Ciência da Computação da Universidade Federal de Santa Maria. Utilizando relatórios do INEP e documentos associados, diagnosticaram que os conteúdos de probabilidade e estatística requerem maior atenção, contradizendo resultados de estudos prévios. O trabalho também destacou problemas na classificação das questões por conteúdos, sugerindo que a adoção de objetos de conhecimento poderia resolver essa limitação.

Com foco no Bacharelado em Ciência da Computação da Universidade Federal de Goiás, Lima et al. (2021), categorizaram as questões do componente específico do ENADE conforme áreas de conhecimento. Essa abordagem identificou as áreas mais influentes para o sucesso no exame e analisou a dificuldade das provas ao longo do tempo, correlacionando-a ao nível de conhecimento esperado dos alunos.

Cunha et al. (2021) desenvolveram um software para analisar os microdados do ENADE de 2005 a 2017, voltado para avaliar o desempenho e identificar defasagens em cursos de ciências da computação. O software revelou uma melhoria significativa no componente específico das provas nesse período, embora o componente geral tenha piorado. O diferencial do trabalho foi a automação do processo analítico, permitindo aplicação em diversos cursos com eficiência.

Vieira et al. (2022) usaram a técnica de agrupamento K-means para minerar microdados do ENADE de 2005 a 2017 em cursos relacionados à computação. O estudo dividiu os estudantes em grupos de desempenho alto, médio e baixo, associando características como escolaridade dos pais, uso de cotas e tipo de escola no ensino médio a esses grupos, trazendo novas perspectivas sobre fatores que influenciam o aprendizado.

Diferentemente dos demais, Brito (2016) focou nos cursos de Administração e utilizou microdados do ENADE 2012 e do Censo da Educação Superior 2012 para investigar a influência do corpo docente no desempenho dos alunos. A pesquisa concluiu que o nível de escolaridade e a quantidade de docentes estão diretamente relacionados ao desempenho discente, questionando a qualidade dos cursos frente à sua expansão.

Esses trabalhos mostram como os microdados e relatórios do INEP podem ser ferramentas para apoiar coordenadores, professores e gestores na melhoria dos processos pedagógicos e da qualidade do ensino. No entanto, nenhum desses estudos utilizou os objetos de conhecimento e a ontologia como ferramenta principal para direcionar ações de forma mais específica nas áreas deficitárias que são as disciplinas, diferencial abordado na proposta deste trabalho.

## 4. Estado da Arte e Originalidade

Análise estatística e avaliação preditiva. a primeira vertente utiliza os microdados do ENADE para realizar análises estatísticas e avaliações de desempenho. Estudos como os de Rendeiro et al. (2023) e Cunha et al. (2021) aplicam análises longitudinais e técnicas estatísticas avançadas para monitorar a evolução e o desempenho dos estudantes em cursos específicos de Tecnologia da Informação. Outros trabalhos, como o de Vieira (2021), exploram o uso de machine learning para prever o desempenho dos alunos ou avaliar a eficácia de políticas públicas. Limitação desta vertente: Embora robustas, estas abordagens tratam os dados predominantemente em seu formato quantitativo (scores e variáveis), sem integrar a semântica do domínio educacional. Eles fornecem resultados agregados (desempenho da turma, previsão de notas), mas não estruturam o conhecimento de forma que permita a inferência automática de competências e lacunas de conhecimento personalizadas.

Arquiteturas para recomendação educacional. a segunda vertente foca na construção de Sistemas de Recomendação (SR) aplicados à educação. Modelos comuns incluem a filtragem colaborativa (FC) ou SRs baseados em conteúdo. Trabalhos nesta área Simon (2018), Campos (2017) e Aguiar (2020), demonstram a eficácia em ambientes onde há alta interação do usuário (avaliações, clicks em materiais didáticos). Limitação desta vertente: Em contextos como o ENADE, a FC é inviável, pois os microdados não contêm o histórico de interação (ratings) dos alunos com materiais didáticos. Portanto, qualquer SR baseado no ENADE deve primeiro resolver a questão de como gerar o perfil de conhecimento (o input de preferência) do aluno antes de aplicar algoritmos de recomendação.

Originalidade e contribuição deste estudo e inserir na intersecção destas vertentes, oferecendo uma solução inovadora para a limitação semântica dos microdados do ENADE. A principal originalidade deste trabalho reside na proposição e construção de uma ontologia como elemento central de pré-processamento semântico. Estruturação semântica, diferentemente dos trabalhos que aplicam análise direta, este artigo utiliza a ontologia para formalizar relações complexas entre cursos, competências, questões e objetos de conhecimento. Isso resolve a limitação da falta de padronização dos relatórios textuais do INEP, transformando-os em conhecimento formal e açãoável (triplas RDF). Geração de conhecimento inferido, a utilização de regras SWRL permite que o sistema inferir novos fatos especificamente a existência de uma KnowledgeGap (Lacuna de conhecimento) a partir dos dados brutos de desempenho. Esta etapa é crucial e não está presente nos estudos correlatos, pois ela cria o perfil semântico de necessidade que alimentará o futuro sistema de recomendação. Alinhamento ao SR, o trabalho estabelece a ponte entre a análise de dados brutos e a personalização de recomendações. Ao gerar a lacuna de conhecimento como um novo fato, o estudo fornece o perfil de necessidade que é a base indispensável para qualquer algoritmo de recomendação em um ambiente de baixo feedback explícito.

Em resumo, enquanto a maioria dos trabalhos utiliza o ENADE como output estatístico, nosso estudo o transforma em input semântico, pavimentando o caminho para o desenvolvimento de sistemas de recomendação educacional verdadeiramente inteligentes e personalizados.

## 5. Metodologia

O desenvolvimento deste projeto envolve uma série de etapas, que vão desde a coleta inicial de dados até a modelagem em uma ontologia<sup>3</sup>. Para isso, trabalhamos com os dados do ENADE 2022 especificamente nos cursos da UFJF que foram avaliados neste ano. Essas etapas podem ser detalhadas como segue, conforme mostrado na figura 2:

**Obtenção dos Microdados do ENADE:** A etapa inicial envolveu o download dos microdados diretamente do site do INEP. Esses dados são disponibilizados em formato ZIP, tornando o processo de aquisição e extração relativamente simples. Os microdados são fundamentais para a análise, pois cada linha da tabela representa um estudante avaliado, contendo variáveis sobre o curso, o ano e o desempenho nas diferentes questões. A ausência de identificadores pessoais, devido à LGPD (Lei Geral de Proteção de Dados Pessoais), exige a criação de um esquema de anonimização, como identificadores do tipo “AlunoX\_Curso\_Ano” (ex.: Aluno1\_1105396\_2022). Essa abordagem assegurou conformidade com a legislação quanto organiza os dados de forma útil para análise.

**Extração de Dados dos Relatórios de Síntese:** Cada relatório contém informações detalhadas sobre as questões do ENADE, divididas em 8 questões gerais (sendo 2 discursivas e 6 objetivas) e 27 questões específicas (sendo 3 discursivas e 24 objetivas). Além disso, para cada questão, é preciso capturar as descrições de competência, perfil e objeto de conhecimento associados. A extração precisa de ferramentas capazes de lidar com PDFs e interpretar corretamente os dados textuais, exigiu técnicas de processamento de linguagem natural (NLP) ou extração programática.

**Filtragem dos Cursos da UFJF Avaliados em 2022:** Realizou-se o download do arquivo ZIP dos microdados do ENADE do ciclo de avaliação de 2022, disponíveis no site do INEP. A análise focou nos cursos da Universidade Federal de Juiz de Fora (UFJF) que participaram da avaliação. No total, foram avaliados 11 cursos, listados a seguir: Administração (22451), Administração Pública (1105396), Ciências Contábeis (1132074), Ciências Econômicas (13090), Direito (Integral e noturno) (13089 / 21860), Jornalismo (Integral e noturno) (21860 / 36356), Psicologia (13113), Serviço Social (13107) e Turismo (21589).

Os cursos foram escolhidos por fazerem parte do mesmo ciclo de avaliação do ano 2022, permitindo a análise da variação de desempenho em um contexto regional controlado e facilitando o mapeamento curricular local. Esse processo envolveu a combinação dos microdados e dos relatórios de síntese, limitando o escopo da análise aos cursos específicos da instituição. Essa etapa foi essencial para evitar ruído nos dados e focar em informações relevantes para a UFJF.

**Construção da Ontologia:** Após consolidar os dados, foi especificada uma ontologia para organizar e relacionar informações como cursos, questões, competências, objetos de conhecimento e desempenho dos estudantes. A ontologia estruturou inicialmente os dados em classes, a saber, Competence, Course, ObjectOfKnowledge, Performance, Profile, e Question (que pode ser subdividida em discursiva e objetiva). Além disso, foram criadas propriedades, objectProperties, como hasCompetence, hasCourse, hasKnowledgeObject, e hasPerformanceQuestion, que conectam os indivíduos na ontologia. Por exemplo, um estudante como “Aluno1\_1105396\_2022” está associado a um curso (“curso1105396\_2022”), a um desempenho (“desempenhoAluno1\_1105396\_2022\_question\_ofg1”) e a questões específicas que avaliam competências e objetos de conhecimento. Essa estrutura possibilita análises semânticas e consultas complexas sobre as relações educacionais.

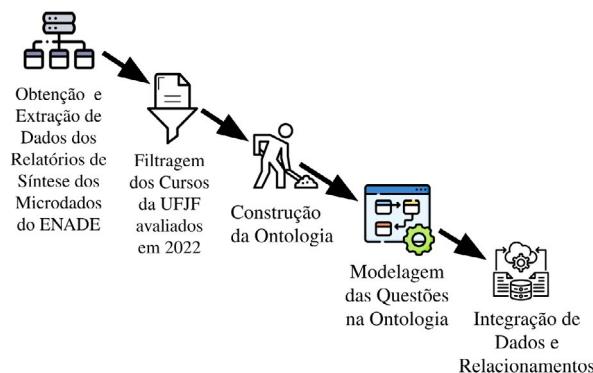
**Modelagem das Questões na Ontologia:** Um desafio específico foi detalhar cada questão na ontologia, vinculando-a às suas respectivas competências, perfis e objetos de conhecimento. Por exemplo, a questão discursiva específica 1 do curso de Administração Pública foi representada como “questao\_dce\_1\_ADMINISTRACAO PUBLICA2022”, com propriedades relacionando-a à competência, ao objeto de conhecimento e ao perfil que avalia. O mesmo processo é replicado para todas as questões discursivas (dce1 a dce3 e dfg1 a dfg2) e objetivas (oce1 a oce27 e ofg1 a ofg8) e assim para todas as questões.

**Integração de Dados e Relacionamentos:** A última etapa foi integrar todas as informações, com os indivíduos e relações corretamente definidos na ontologia. Essa integração permite que os dados sejam consultados e analisados de maneira semântica, promovendo insights sobre o desempenho acadêmico e as lacunas de aprendizado nos cursos avaliados.

Esses desafios demandaram um planejamento e a aplicação das técnicas especializadas de extração, filtragem, modelagem e análise de dados. Apesar das dificuldades, o resultado foi uma solução inicial da ontologia que permite

entender as deficiências de aprendizado, contribuindo, quando a solução proposta estiver consolidada, para a melhoria da educação superior na UFJF e podendo servir de base para iniciativas semelhantes em outras instituições.

Figura 2. Metodologia

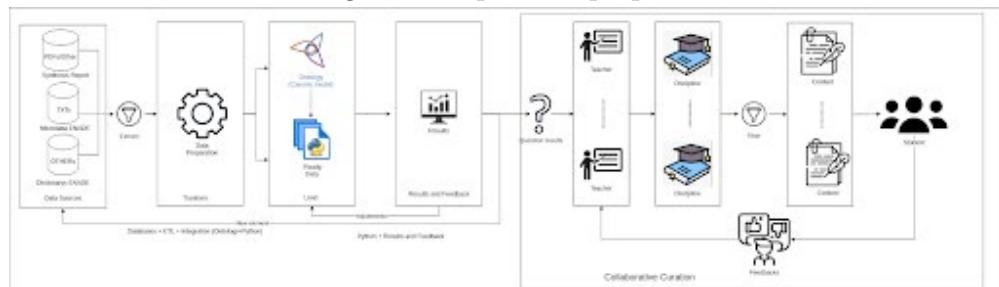


Fonte: Elaborado pelos autores (2025)

## 6. Arquitetura

A arquitetura conceitual proposta é composta pelo modelo ontológico, que organiza os dados educacionais, além de um sistema de recomendação que utiliza essa ontologia para sugerir conteúdos educativos relevantes. Conforme é apresentado na figura 3 a arquitetura proposta integra diversas fontes de dados, como PDFs, relatórios de síntese e os microdados do ENADE, provenientes de diferentes sistemas do INEP. O fluxo de dados começa nas fontes, representadas pelos arquivos .pdf, .txt e outros formatos de relatório. Detalhamos a seguir os principais componentes:

Figura 3. Arquitetura proposta



Fonte: Elaborado pelos autores (2025)

**Data Sources (Fontes de Dados):** As fontes de dados incluem: PDFs/OUTROS: Datasets dos relatórios de síntese de cada curso avaliado pelo INEP no ENADE. .txts: Microdados. OUTROS: Relatórios de síntese. Extração, Transformação e Carga (ETL): O processo de ETL (Extract, Transform, Load) é responsável por extrair, transformar e carregar os dados para o ambiente de análise. A ontologia, utilizando uma estrutura semântica, organiza e estrutura os dados extraídos, permitindo sua manipulação por meio do Python, que lê e processa essas informações para análise.

**Análise de Dados:** Após a carga dos dados, o sistema realiza a análise com o processamento de inferências na ontologia e uso de algoritmos de aprendizado de máquina (machine learning) para extrair padrões e previsões relevantes para as questões de desempenho dos alunos.

**Resultados e Feedback:** Os resultados da análise alimentam um ciclo de feedback que será essencial para otimizar os processos de ensino e ajustar as práticas pedagógicas, baseando-se nas previsões e insights gerados pela análise de dados. Já a colaboração e Feedback complementa a arquitetura, mostrando como os resultados das análises impactam o processo de ensino e a colaboração entre os envolvidos. Assim, a arquitetura apoia diretamente o(a) professor(a) com dados sobre o desempenho dos alunos nas questões do ENADE, permitindo uma visão clara das áreas que precisam de mais atenção nas disciplinas. O feedback é utilizado para ajustar o conteúdo e os métodos de ensino.

Com base nos resultados da análise, é aplicado um filtro para identificar quais conteúdos precisam ser reforçados ou revisados, permitindo que os professores adaptem as aulas de acordo com as necessidades específicas dos alunos. O feedback gerado pelos resultados da análise também pode auxiliar os estudantes, possibilitando que eles identifiquem suas próprias lacunas de conhecimento e recebam recomendações personalizadas de estudo. A colaboração entre professores, alunos e sistemas de recomendação é facilitada por um ciclo contínuo de análise e feedback, criando um ambiente educacional colaborativo e dinâmico. A interação e cooperação entre todos os envolvidos promove a melhoria contínua do processo de ensino-aprendizagem.

## 7. Desenvolvimento da Solução e Tecnologias utilizadas

Para verificar a viabilidade da proposta inicial de uso da ontologia, detalhamos a seguir questões tecnológicas relacionadas à implementação da arquitetura:

**Coleta e Filtragem dos Microdados:** Realizou-se o download do arquivo ZIP dos microdados do ENADE 2022, com foco no arquivo `microdados2022_arq3`. Esse arquivo inclui informações detalhadas sobre edição, código de curso, itens válidos por seção da prova, vetores, tipos de presença, tipos de situação das questões discursivas, notas de Formação Geral e Componente Específico, e outras variáveis fundamentais para a análise, conforme mostrado na figura 4, como:

Identificadores gerais: NU\_ANO, CO\_CURSO.

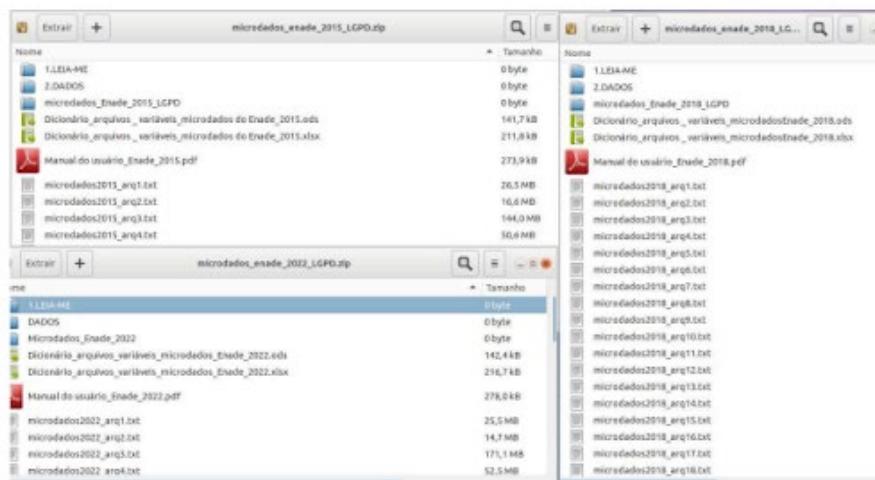
Vetores e gabaritos: DS\_VT\_GAB\_OFG\_FIN, DS\_VT\_GAB\_OCE\_FIN.

Notas: NT\_OBJ\_FG, NT\_DIS\_FG, NT\_OBJ\_CE, NT\_DIS\_CE.

Itens específicos e respostas: NU\_ITEM\_OFG, DS\_VT\_ESC\_OFG, DS\_VT\_ACE\_OCE.

Com auxílio de ferramentas como Excel e LibreOffice, aplicou-se um filtro para selecionar os cursos da UFJF avaliados no ENADE 2022. O resultado foi salvo em arquivos de saída nos formatos .xlsx e .csv, denominados `cursos_ufjf_enade2022.xlsx` e `cursos_ufjf_enade2022.csv`.

Figura 4. Microdados ENADE 2022



Fonte: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>

**Fonte dos Dados:** Para a análise e predição das notas do ENADE, foram utilizados os microdados do ano de 2022. Esses dados estão disponíveis em arquivos de texto (.TXT).

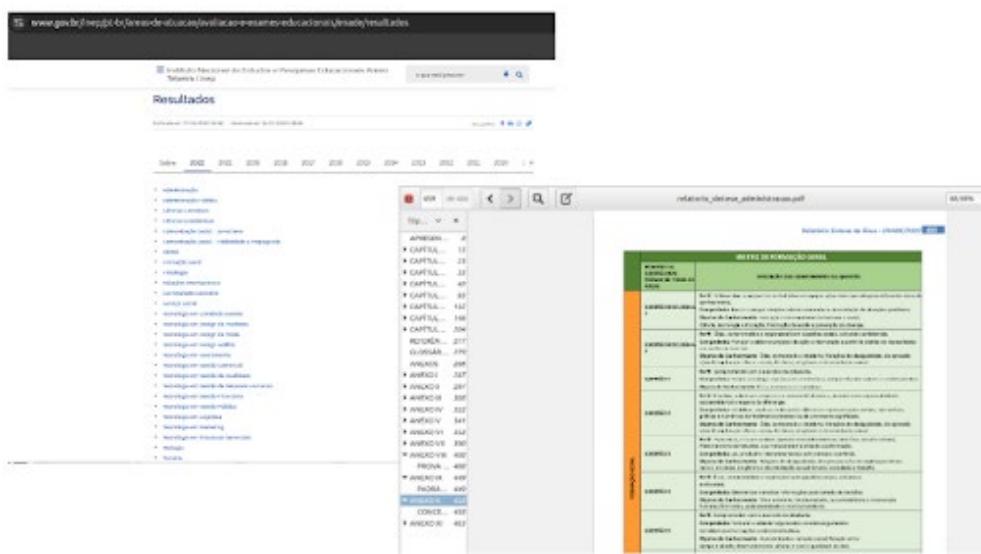
**Microdados ENADE 2022.TXTs:** Os microdados do ENADE referentes ao ano de 2022, conforme o ciclo Avaliativo Trienal estabelecido na Portaria Normativa MEC nº 840/2018. Este ciclo inclui avaliações anuais de diversos cursos de graduação. Em 2022 e ano III - Bacharelado em: Ciências Sociais e Áreas afins; Ciências Humanas e áreas afins que não tenham cursos também avaliados no âmbito da licenciatura;

**Cursos Avaliados:** os cursos da Universidade Federal de Juiz de Fora (UFJF) que participaram das avaliações. No total, foram avaliados 11 cursos: Administração (22451); Administração Pública (1105396); Ciências Contábeis

(1132074); Ciências Econômicas (13090); Direito (Integral e noturno) (13089 / 21860); Jornalismo (Integral e noturno) (21860 /36356); Psicologia (13113); Serviço Social (13107) e Turismo (21589)

**Extração dos Relatórios de Síntese:** Utilizando Python, foi automatizada a extração de informações dos relatórios de síntese disponíveis no site oficial do ENADE <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/resultados>, conforme e mostrado na figura 5 um exemplo do curso de Administração Pública. Para os cursos selecionados da UFJF, identificaram-se: A matriz de formação geral das questões. A identificação de cada questão (número e tipo). A descrição dos componentes associados, como competências, perfis e objetos de conhecimento.

Figura 5. Relatório de Síntese ENADE 2022 com exemplo do curso Administração Pública



Fonte:<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>

As informações extraídas foram organizadas em arquivos .xlsx e .csv, categorizados por curso, para facilitar o cruzamento e a análise posterior fazendo python, conforme a imagem do código na figura 6. Foram verificados diversos erros de português, falta de vírgulas e outros correspondentes. Como essas ocorrências não tem um padrão muito bem definido, isso prejudicou muito o desenvolvimento e a padronização para extração dos pdfs.

Figura 6. Código de extração de Relatório de Síntese ENADE 2022

```
extratorFull.py x
1 import os
2 import glob
3 import csv
4 import re
5 import pandas as pd # Importando a biblioteca pandas
6
7 # Cuidado para a pasta que contém os arquivos CSV
8 input_folder_path = 'CSVs/'
9 output_folder_path = 'filtrados/'
10
11 # Criar a pasta 'filtrados' se não existir
12 os.makedirs(output_folder_path, exist_ok=True)
13
14 # Pedro para encontrar todos os arquivos CSV que começam com 'conteúdo prova' e terminam com '2022.csv'
15 csv_files = glob.glob(os.path.join(input_folder_path, 'conteúdo prova *2022.csv'))
16
17 # Itera sobre cada arquivo CSV encontrado
18 for csv_file in csv_files:
19     # Inicializa listas para armazenar os dados das colunas
20     col0 = []
21     col1 = []
22     col2 = []
23
24     # Lê o arquivo CSV
25     with open(csv_file, 'r', newline='', encoding='utf-8') as file:
26         reader = csv.reader(file)
27
28         # Itera sobre as linhas do arquivo
29         for row in reader:
30             # Ignora a linha se esta vazia e tem pelo menos 3 colunas
31             if row and len(row) >= 3:
32                 col0.append(row[0]) # Adiciona o valor da primeira coluna
33                 col1.append(row[1]) # Adiciona o valor da segunda coluna
34                 col2.append(row[2]) # Adiciona o valor da terceira coluna
35
36     # Exibe os dados coletados (opcional)
37     print(f'Processando arquivo: {os.path.basename(csv_file)}')
38
39     # Definindo o padrão para as questões válidas
40     pattern = re.compile(r'^[Q]UESTÃO [1-5] | [Q]UESTÃO [1-9] | [Q]UESTÃO [1-2][0-9] | [Q]UESTÃO [3-5][0-5]$')
41
42     # Listas para armazenar os dados que serão salvos no CSV
43     questoes_validas = []
44     perfis = []
45     competencias = []
46     objetos_conhecimento = []
47
48
49 # Filtrando os dados e dividindo as strings em col2
50 for i in range(len(col1)):
51     if pattern.match(col1[i]): # Verifica se a string corresponde ao padrão
52         questoes_validas.append(col1[i]) # Adiciona a questão válida
53
54     # Extrai o texto de coluna
55     partes = re.split(r'\s+|,|;|&|:|<|>', col2[i]) # Remove espaços em branco, vírgulas, ponto e vírgula, parênteses e hífens
56     partes = [parte.strip() for parte in partes] # Remove espaços em branco unnecessary
57
58     perfil = ''
59     competencia = ''
60     objeto_conhecimento = ''
61
62     # Extraiendo PERFIL, COMPETÊNCIA e OBJETO DE CONHECIMENTO
63     for parte in partes:
64         if parte.startswith('Perfíl'):
65             perfil += parte.split(':')[1][1].strip() if len(parte.split(':')) > 1 else '' # Extrai o texto após 'Perfíl'
66         elif parte.startswith('Competência'):
67             competencia += parte.split(':')[1][1].strip() if len(parte.split(':')) > 1 else '' # Extrai o texto após 'Competência'
68         else:
69             objeto_comhecimento = parte.replace('Objeto de Conhecimento', '') # Remove o nome do objeto de conhecimento
70             objeto_comhecimento += objeto_comhecimento + ' ' # Extrai e concatena o texto após ' '
71
72     # Adicionando os dados às listas correspondentes
73     perfis.append(perfil)
74     competencias.append(competencia)
75     objetos_conhecimento.append(objeto_comhecimento.strip())
76
77 # Criação de Dataframe pandas para armazenar os dados processados
78 df = pd.DataFrame({
79     'QUESTÃO': questoes_validas,
80     'PERFIL': perfis,
81     'COMPETÊNCIA': competencias,
82     'OBJETO DE CONHECIMENTO': objetos_conhecimento,
83 })
84
85 # Definindo o nome do arquivo de saída baseado no nome do arquivo original
86 output_file_name = f'questoes_{os.path.basename(csv_file)}'
87 csv_output_path_final = os.path.join(output_folder_path, output_file_name)
88
89 # Salvando o Dataframe em um novo arquivo CSV na pasta 'filtrados'
90 df.to_csv(csv_output_path_final, index=False, encoding='utf-8')
91
92 print(f'Tabela salva como: {output_file_name}')
93
94 print('Todos os arquivos foram processados e salvos na pasta "filtrados".')
```

Fonte: Elaborado pelos autores (2025)

Ao final, foram gerados os seguintes arquivos, conforme a figura 7, contendo a matriz de conteúdo de cada curso, no formato em csv. Preparação para a Elaboração da Ontologia: Com os dados dos microdados filtrados e as descrições das questões extraídas dos relatórios, o próximo passo foi estruturar essas informações em uma

ontologia. Essa etapa combina as variáveis selecionadas e a descrição das questões para criar uma representação semântica das relações entre alunos, cursos, competências e objetos de conhecimento, utilizando ferramentas como Protégé e formatos como OWL. Foi criado uma ontologia padrão, detalhada nas figuras 7,8,9 e 10. Destaca-se nas figuras as classes Competence, Course, ObjectOfKnowledge, Performance, Profile, e Question (que pode ser subdividida em discursiva e objetiva) e as object properties hasCompetence, hasCourse, hasKnowledgeObject, e hasPerformanceQuestion, que conectam os indivíduos na ontologia.

Figura 7. Código de extração de Relatório de Síntese ENADE 2022

Fonte: Elaborado pelos autores (2025)

Figura 8. Construção da ontologia - Classes e object property e Data property



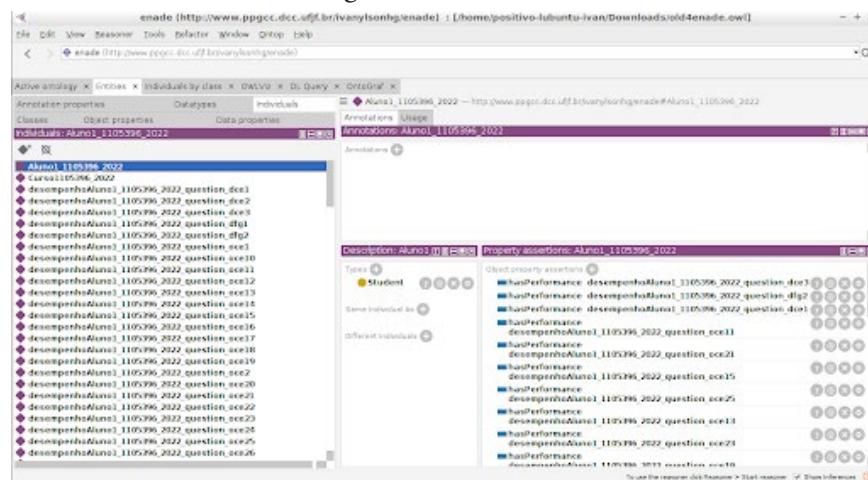
Fonte: Elaborado pelos autores (2025)

Na figura 9, vale destacar que durante a elaboração da ontologia, adotou-se um padrão estruturado para a nomeação dos indivíduos, garantindo clareza e consistência na representação semântica dos dados. Esse padrão reflete as relações entre alunos, cursos, desempenho em questões e os atributos associados, como competência, objeto de conhecimento e perfil. Para cada questão, há a definição dos seus componentes, como: Competências associadas:

Exemplo: questao\_dce\_1\_ADMINISTRAÇÃO PÚBLICA2022\_competencia. Objeto de conhecimento: Exemplo: questao\_dce\_1\_ADMINISTRAÇÃO PÚBLICA2022\_objetoConhecimento. Perfil da questão: Exemplo: questao\_dce\_1\_ADMINISTRAÇÃO PÚBLICA2022\_perfil. Esse padrão é replicado para todas as questões, organizando as relações em categorias discursivas gerais (DFG), discursivas específicas (DCE), objetivas gerais (OFG) e objetivas específicas (OCE). Cada instância também é associada ao curso correspondente (por exemplo,

Administração Pública em 2022) e suas propriedades semânticas por meio de relações como hasCompetence, hasKnowledgeObject, hasPerformanceQuestion, e hasProfile.

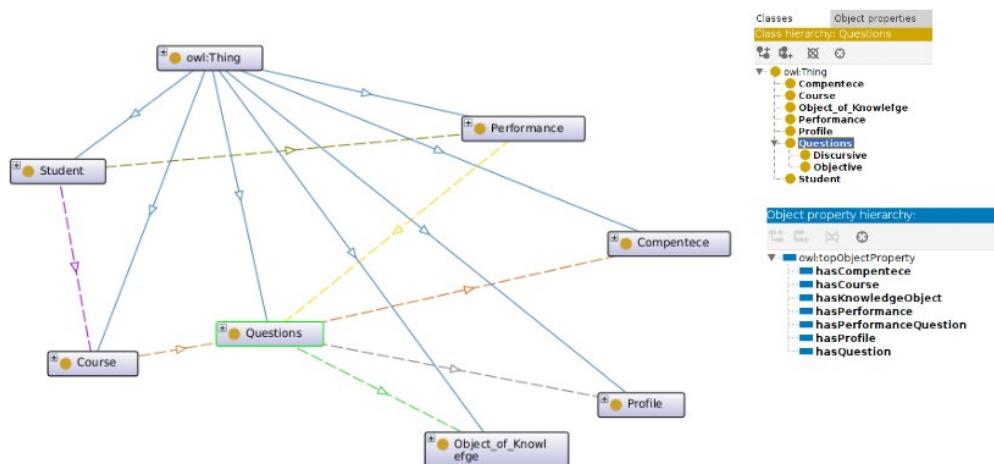
Figura 9. Indivíduos



Fonte: Elaborado pelos autores (2025)

Essa estrutura padronizada facilita a criação e manutenção da ontologia, promovendo uma organização clara para análises posteriores. O modelo mostrado na figura 10 facilita a consulta, rastreabilidade e interpretação dos dados avaliativos, permitindo análises detalhadas sobre alunos, cursos e questões, além de assegurar um vínculo lógico entre os elementos por meio de propriedades semânticas.

Figura 10. Ontologia

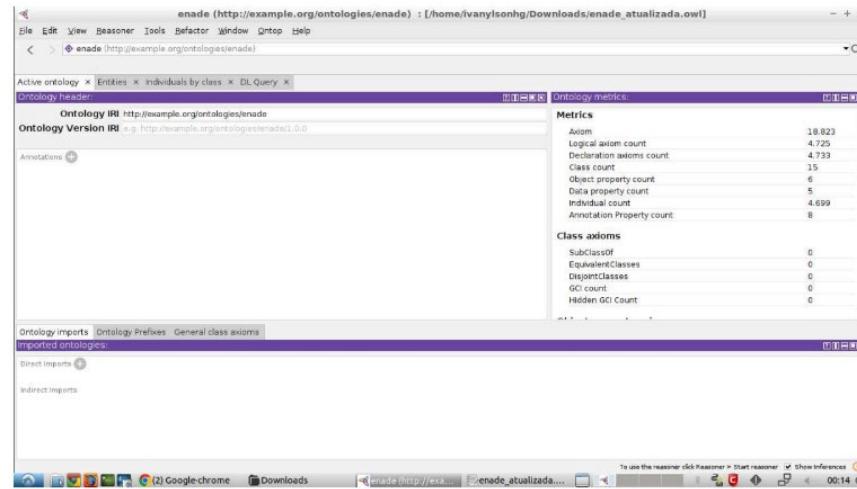


Fonte: Elaborado pelos autores (2025)

Para viabilizar o processamento inteligente e a inferência de conhecimento, a ontologia deve ser rica axiomaticamente. Utilizamos as Regras SWRL (Semantic Web Rule Language), que são fundamentais para expressar lógica complexa sobre a ontologia. O SWRL permite criar novos relacionamentos (triplas RDF) a partir de dados existentes, combinando classes, propriedades e literais.

Definimos um conjunto de regras SWRL para inferir novo conhecimento no grafo RDF, focando no desempenho dos alunos, nos relacionamentos entre cursos e nas competências ligadas a questões do ENADE. A aplicação dessas regras transforma dados brutos em informação inferida e açãoável, sendo um passo crucial para a análise semântica e para a identificação de lacunas de aprendizado. A ontologia, conforme a figura 11, mostra as métricas da ontologia atualizada.

Figura 11. Ontologia - enade\_atualizada



Fonte: Elaborado pelos autores (2025)

## 8. Demontração e Avaliação Funcional da Ontologia

O principal objetivo deste trabalho é criar uma arquitetura semântica para processar os microdados do ENADE e gerar conhecimento inferido sobre lacunas de aprendizado. Para validar essa abordagem, conduzimos uma Prova de Conceito (PoC) focada estritamente na capacidade de inferência da ontologia. A PoC não é um sistema de recomendação completo (trabalho futuro), mas sim uma demonstração técnica de que: A ontologia, junto com as regras SWRL, consegue transformar dados brutos de desempenho em informação acionável e estruturada (lacunas de conhecimento). Isso comprova a viabilidade técnica da arquitetura. A metodologia da PoC seguiu três etapas: População da ontologia; Execução das regras de inferência (usando HermiT/SWRL) e Análise dos novos relacionamentos inferidos. Esta seção atesta a viabilidade técnica, demonstrando a capacidade da ontologia de recuperar (via SPARQL) e inferir dados complexos. Todas as consultas utilizam o prefixo definitivo. PREFIX: <<http://www.ppgcc.dcc.ufjf.br/ivanylsonhg/enade#>>. Consultas de Exploração e Busca de Dados (Pré-Inferência) Estas consultas demonstram a capacidade de recuperação de informações específicas da ontologia. Detalhes de uma questão específica (Estrutura da ontologia). Recupera id, competência, objeto de conhecimento e perfil de uma questão, verificando a estrutura de triplas aninhadas, conforme mostrado na figura 12.

Figura 12. Detalhes de uma Questão Específica

ID	Competence	ObjetoConhecimento	Perfil
questao_ofg_6_ADMINISTRAÇÃO PÚBLICA2022	"Identificar, compreender e analisar situações-problema"	"Processos de globalização e política internacional; Técnico-Humanista, critica e sensível, apoiado em conhecimento"	"Humanista, critica e sensível, apoiado em conhecimento"

Fonte: Elaborado pelos autores (2025)

Busca textual por palavra-chave (Capacidade de busca semântica). Recupera questões com base em termos específicos (globalização ou Territórios) dentro da descrição do objeto de conhecimento, demonstrando a capacidade de busca textual avançada, conforme descrito na figura 13.

Figura 13. Busca Textual por Palavra-Chave

The screenshot shows the Protege-5.6.7 interface with a SPARQL query results table. The query retrieves questions related to 'globalização' and 'Territórios'. The results table has two columns: 'QuestaoURI' and 'ObjetoConhecimento'. The first row contains 'questao\_ofg\_4\_ADMINISTRAÇÃO PÚBLICA2022' and 'Meio ambiente: biodiversidade, sustentabilidade e intervenção humana; Territórios, sociodiversidade e mu...'. The second row contains 'questao\_ofg\_6\_ADMINISTRAÇÃO PÚBLICA2022' and 'Processos de globalização e política internacional; Territórios, sociodiversidade e multiculturalismo.'.

QuestaoURI	ObjetoConhecimento
questao_ofg_4_ADMINISTRAÇÃO PÚBLICA2022	"Meio ambiente: biodiversidade, sustentabilidade e intervenção humana; Territórios, sociodiversidade e mu..."
questao_ofg_6_ADMINISTRAÇÃO PÚBLICA2022	"Processos de globalização e política internacional; Territórios, sociodiversidade e multiculturalismo."

Fonte: Elaborado pelos autores (2025)

Listagem de Desempenho Completo do Aluno 1. Demonstra o estado inicial dos dados de desempenho (notas) para o aluno alvo, conforme é mostrado na figura 14.

Figura 14. Listagem de Desempenho

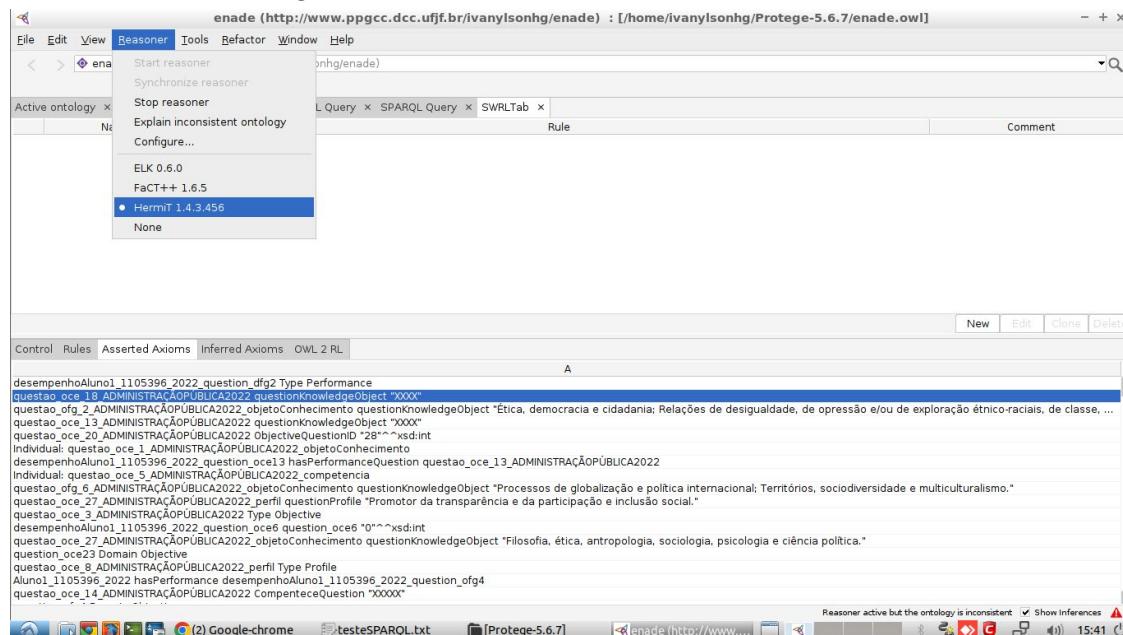
The screenshot shows the Protege-5.6.7 interface with a SPARQL query results table. The query retrieves distinct student information along with their performance data. The results table has six columns: 'Aluno', 'Questao', 'TipoNota', 'Nota', 'Curso', and 'Ano'. The table lists various students (Aluno1\_1105396\_2022, Aluno1\_1105396\_2022, etc.) and their corresponding performance data across different subjects and years.

Aluno	Questao	TipoNota	Nota	Curso	Ano
Aluno1_1105396_2022	questao_dce_1_ADMINISTRAÇÃO PÚBLICA question_dce1	"0"^^<http://www.w3.org/201/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_dce_2_ADMINISTRAÇÃO PÚBLICA question_dce2	"0"^^<http://www.w3.org/201/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_dce_3_ADMINISTRAÇÃO PÚBLICA question_dce3	"0"^^<http://www.w3.org/201/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_dfg_1_ADMINISTRAÇÃO PÚBLICA question_dfg1	"0"^^<http://www.w3.org/201/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_dfg_2_ADMINISTRAÇÃO PÚBLICA question_dfg2	"0"^^<http://www.w3.org/201/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_10_ADMINISTRAÇÃO PÚBLICA question_oce10	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_11_ADMINISTRAÇÃO PÚBLICA question_oce11	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_12_ADMINISTRAÇÃO PÚBLICA question_oce12	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_13_ADMINISTRAÇÃO PÚBLICA question_oce13	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_14_ADMINISTRAÇÃO PÚBLICA question_oce14	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_15_ADMINISTRAÇÃO PÚBLICA question_oce15	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_16_ADMINISTRAÇÃO PÚBLICA question_oce16	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_17_ADMINISTRAÇÃO PÚBLICA question_oce17	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			
Aluno1_1105396_2022	questao_oce_18_ADMINISTRAÇÃO PÚBLICA question_oce18	"0"^^<http://www.w3.org/2001/XMLSchema#1105396>^^<http://www.w3.org/2022/XMLSchema#2022>^^<http://www.w3.org/2001/XMLSchema#			

Fonte: Elaborado pelos autores (2025)

Geração de conhecimento (Inferência por HermiT), a principal funcionalidade da PoC é demonstrar a capacidade de gerar um diagnóstico de alto nível, conforme mostrado na figura 15. Isso é feito por meio da regra de inferência SWRL, processada pelo HermiT. Cenário e o diagnóstico de dificuldade, a regra SWRL ativa, a regra transforma o erro em uma questão (:hasDifficultyWith) em uma dificuldade no conceito principal associado à questão (:hasDifficultyWithKnowledgeObject).

Figura 15. Reasoner usando a inferência Hermit



Fonte: Elaborado pelos autores (2025)

## 9. Discussão

A ontologia demonstrou ser uma solução robusta para estruturar os microdados brutos do ENADE, superando a natureza despadronizada dos relatórios originais. A Prova de Conceito (PoC) confirmou que a abordagem semântica é capaz de gerar conhecimento inferido (lacunas de aprendizado), que serve como o input semântico ideal para a futura camada de recomendação, facilitando a integração com machine learning (ML). A arquitetura proposta aprimora a análise dos dados ao possibilitar sua organização semântica e hierárquica, preparando-os para o processamento avançado. Contudo, a implementação prática desta arquitetura em ambientes educacionais reais exige considerações adicionais. Será necessária a validação por profissionais de pedagogia e tecnologia, além da adaptação às particularidades de cada instituição. Desafios de escalabilidade, integração com plataformas existentes e treinamento de usuários devem ser gerenciados para garantir uma implantação sustentável. O avanço desta proposta dependerá do desenvolvimento de estratégias de suporte técnico, validação e adaptação. A integração futura com ML promete potencializar as recomendações, permitindo intervenções pedagógicas mais direcionadas e efetivas.

## 10. Considerações Finais

Este artigo estabeleceu uma arquitetura baseada em ontologia como alicerce para um futuro sistema de recomendação de conteúdo educacional, utilizando os microdados do ENADE. Detalhamos a construção e validação da ontologia, confirmando sua capacidade de estruturar os dados de forma semântica, o que promove uma compreensão aprofundada do desempenho e das lacunas de aprendizagem dos estudantes. Esta pesquisa, portanto, avança a análise de dados educacionais e a personalização do ensino superior. Como trabalhos futuros, planejamos: desenvolver o sistema de recomendação, implementando filtragem colaborativa e machine learning para gerar sugestões personalizadas a partir das lacunas inferidas; realizar a validação empírica com docentes em ambientes reais, a fim de verificar a utilidade prática das sugestões e desenvolver estratégias pedagógicas mais eficazes; e otimizar o processo ETL para mitigar a baixa padronização dos relatórios do ENADE, garantindo a qualidade e a atualização contínua dos dados.

## Referências

AGUIAR, Janderson Jason Barbosa; DE ARAÚJO, Joseana Macêdo Fechine Régis; DE BARROS COSTA, Evandro. Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem. *Brazilian Journal of Development*, v. 6, n. 1, p. 273-286, 2020.

BELO, Orlando. OntoSINAES: Uma Rede de Ontologias para o Sistema Nacional de Avaliação da Educação Superior no Brasil. *OntoSINAES: An Ontology Network for the National Evaluation System of Higher Education in Brazil*. 2017. Acesso: 14 Ago 2025. Disponível em: (<https://repositorium.sdum.uminho.pt/bitstream/1822/71920/1/2016-CN-CAPSI-Silva%26Belo-CRP.pdf>).

BERNERS-LEE, T. MILLER, E. The Semantic Web Lifts Off. 2002, Acesso: 14 Ago, 2025 (Disponível em: <http://www.ercim.org/publication/Ercim\ News/enw51/berners-lee.html>).

CAMPOS, Aline de, et al. “Mapeamento de soluções tecnológicas em sistemas de recomendação educacionais em âmbito brasileiro.” *Informativa na educação: teoria & prática*. Porto Alegre. Vol. 20, n. 3 (set./dez. 2017), p. 78-93 (2017).

CHOJI, Mayk Fernando et al. Mineração de dados do Enade de 2016 a 2018: uma análise sobre o município de Araçatuba/SP. *RENOTE-Revista Novas Tecnologias na Educação*, v. 19, n. 2, p. 183-192, 2020 2 1 . Acesso: 14 Ago 2025. Disponível em: (<https://repositorio.usp.br/bitstreams/8288fcf3-13c9-41b0-bde7-46109446de79>).

CRUZ, Breno Brito; MACIEL, Kilbert Amorim. SAEN: Um sistema de análise de dados baseado em resultados do ENADE. In: *ESCOLA REGIONAL DE COMPUTAÇÃO DO CEARÁ, MARANHÃO E PIAUÍ (ERCEMAPI)*, 11., 2023, Aracati/CE. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 92-101. Acesso: 14 Ago 2025. Disponível em: (<https://doi.org/10.5753/ercemapi.2023.236410>).

CUNHA, Renan; SALES, Cláudomiro; SANTOS, Reginaldo. Análise automática com os microdados do enade para melhoria do ensino dos cursos de ciência da computação. In: *Workshop sobre Educação em Computação (WEI)*. SBC, 2021. p. 208-217. Acesso: 12 Ago 2025. Disponível em: (<https://sol.sbc.org.br/index.php/wei/article/view/15912>).

DE LIMA, Júnio César; DE CARVALHO, Cedric L. Uma Visão da Web Semântica. Technical Report INF\_001/94, Instituto de Informática-Universidade Federal de Goiás, 2004. Acesso: 12 Ago 2025. Disponível em: ([https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-04.pdf](https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04.pdf)).

DE OLIVEIRA, J. J. Web Semântica. 2003. Acesso: 12 Ago 2025. Disponível em: (<http://limbo.ime.usp.br/mac339/index.php/IcicWebSemantica>).

DE OLIVEIRA, Marcella N. et al. SIMULA-ENADE: Support System for University Preparation for the National Student Performance Examination. In: *Anais do XXXIV Simpósio Brasileiro de Informativa na Educação*. SBC, 2023. p. 36-46. Acesso: 14 Ago 2025. Disponível em: (<https://sol.sbc.org.br/index.php/sbie/article/view/26648>).

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. Web Semântica. Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. 18, p. 20-39, 2004. Acesso: 14 Ago 2025. Disponível em ([http://www.encontros-bibli.ufsc.br/Edicao\\_18/2\\_Web\\_Semantica.pdf](http://www.encontros-bibli.ufsc.br/Edicao_18/2_Web_Semantica.pdf)).

ENADE (2022). Resultados do Enade 2022 e Perspectivas para a Avaliação da Educação Superior a partir de 2025 Acesso em: 25 Ago. 2025. Disponível em: ([https://download.inep.gov.br/educacao\\_superior/enade/apresentacao/2022/apresentacao\\_resultados\\_e\\_nade\\_2022.pdf](https://download.inep.gov.br/educacao_superior/enade/apresentacao/2022/apresentacao_resultados_e_nade_2022.pdf)).

FERREIRA, J. S. P. Elementos básicos de aprendizagem de máquina. *Intermaths*, 4(2), 54-74. 2023. <https://doi.org/10.22481/intermaths.v4i2.13422>. Acesso em: 25 Ago. 2025. Disponível em: (<https://periodicos2.uesb.br/index.php/intermaths/article/view/13422/8332>).

GOV.BR. Exame Nacional de Desempenho dos Estudantes (Enade). Brasília, DF, 2025. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>. Acesso em: 12 dez. 2025.

INEP. Manual Enade 2016. Brasília, DF, 2016. Disponível em: [https://download.inep.gov.br/educacao\\_superior/enade/manuais/manual\\_do\\_enade\\_28092016.pdf](https://download.inep.gov.br/educacao_superior/enade/manuais/manual_do_enade_28092016.pdf). Acesso em: 12 dez. 2025.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Exame nacional de desempenho dos estudantes - ENADE. Acesso em: 18 Ago. 2025. Disponível em: (<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>).

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados Enade 2015. Brasília: Inep, 2016. Disponível em (<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>). Acesso em: 13 Ago. 2025.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados Enade 2018. Brasília: Inep, 2022. Disponível em: (<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>). Acesso em: 15 Ago. 2025.

LIMA, P. da S. N.; AMBRÓSIO, A. P. L.; OLIVEIRA, J. L. dos S.; CARVALHO, C. L. de. Análise de conteúdo das provas do Enade para os alunos do curso de Bacharelado em Ciência da Computação . Revista Brasileira de Informática na Educação, [S. l.], v. 29, p. 385–413, 2021. DOI: 10.5753/rbie.2021.29.0.385. Acesso em: 15 Ago. 2025. Disponível em: (<https://journals-sol.sbc.org.br/index.php/rbie/article/view/2981>).

OLIVEIRA, Michael Uewerton Targino de. Análise sobre dados do enade do curso de sistemas da informação da UFPB do campus de Rio Tinto. Trabalho de conclusão de curso, 2022. Acesso: 28 Ago 2025. Disponível em: (<https://repositorio.ufpb.br/jspui/handle/123456789/29088>).

OLIVEIRA, Samuel. Classificação automática de questões do ENADE utilizando o algoritmo KNN. 2021. Acesso: 15 Ago 2025. Disponível em: (<https://dspace.mackenzie.br/handle/10899/31047>).

PICKLER, Maria Elisa Valentim. Web Semântica: ontologias como ferramentas de representação do conhecimento. Perspectivas em Ciência da Informação, v. 12, p. 65-83, 2007. Acesso: 15 Ago 2025. Disponível em: (<https://www.scielo.br/j/pci/a/HHdw6KMPG45HxwShcwTmFSs/>).

PORTAL BRASILEIRO DE DADOS ABERTOS. Sobre. [S.l.]: [s.n.], [2025]. Disponível em: <https://dados.gov.br/dados/conteudo/sobre>. Acesso em: 12 dez. 2025.

SIMON, Augusto. “Ontologias na educação: uma proposta de aplicação em sistemas de recomendação.” (2018). Acesso: 20 Dez 2025. Disponível em: (<https://lume.ufrgs.br/handle/10183/220440>).

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. Ci. Inf., Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004. Acesso: 15 Ago 2025. Disponível em: (<https://www.scielo.br/j/ci/a/sp3XpmZhXw384H5Fw9H89YL/>).

RENDEIRO, Pedro et al. Avaliação de Desempenho em Cursos de Engenharia de Computação Baseada nos Microdados do ENADE. In: Anais do XXXI Workshop sobre Educação em Computação. SBC, 2023. p. 339-349. Acesso: 15 Ago 2025 Disponível em: (<https://sol.sbc.org.br/index.php/wei/article/download/24915/24736/>).

ROSA, Eliene Ribeiro et al. Estudo Exploratório através de Análises Longitudinais aplicado à Ciência da Computação a partir da Base de Dados do ENADE. Revista Brasileira de Informática na Educação, v. 29, p. 1463-1486, 2021. Acesso: 16 Ago 2025. Disponível em: (<https://journals-sol.sbc.org.br/index.php/rbie/article/download/2073/1890>).

VIEIRA, Gutemberg Assunção. Uso de aprendizado de máquina na avaliação de políticas públicas: uma revisão de escopo. 2023. Acesso: 26 Ago 2025 Disponível em: (<https://repositorio.enap.gov.br/bitstream/1/7784/1/Disserta%C3%A7%C3%A3o%20-%20Gutemberg%20Assun%C3%A7%C3%A3o%20vers%C3%A3o%20com%20ficha.pdf>).

W3C. W3C Semantic Web Activity. 2001. Acesso: 15 Ago 2025 Disponível em: (<http://www.w3.org/2001/12/semweb-fin/w3csw>).

W3C. W3C PROV-Overview. 2013. Acesso: 15 Ago 2025 Disponível em: (<https://www.w3.org/TR/prov-overview/>)