

Abordagem para Análise de Múltiplas Fontes de Dados de Evasão Escolar

Eduardo Moura da Silva
Programa de Pós-graduação em
Computação Aplicada
Instituto Federal do Espírito Santo
Serra, ES, Brasil
eduardo.silva@ifes.edu.br

Fabiano Borges Ruy
Programa de Pós-graduação em
Computação Aplicada
Instituto Federal do Espírito Santo
Serra, ES, Brasil
fabianoruy@ifes.edu.br

Filipe Wall Mutz
Programa de Pós-graduação em
Computação Aplicada
Instituto Federal do Espírito Santo
Serra, ES, Brasil
filipe.mutz@ifes.edu.br

ABSTRACT

Scholar dropout is a challenge for educational institutions. It is caused by many factors that should be known to be treated. Some analytical tools can help in this intent, especially when the proper data is well organized in a consistent structure and format. This paper presents a technical approach to allow homogeneous analysis for different scholar data sources, applying it to the dropout domain. This proposal makes use of an ontology to provide a better understanding of the domain semantics, and to act as an interlanguage for the creation of standardized repositories from where dropout data can be analyzed.

KEYWORDS

Evasão Escolar, Integração Semântica, Ontologias, Repositório de Dados, Análise de Dados

1 INTRODUÇÃO

A evasão escolar é uma questão que preocupa não apenas os gestores escolares, mas também o Estado e a sociedade em geral, pois é um fenômeno que gera impacto para todos. Segundo [1], esse é um problema que preocupa as instituições de ensino públicas e privadas, pois as saídas de alunos provocam graves consequências sociais, acadêmicas e econômicas. Com isso, ao longo do tempo, diversas pesquisas vêm sendo realizadas para melhor compreender esse fato que ocorre frequentemente nas instituições de ensino. Em [2], os autores apontam que, para a instituição, a evasão acarreta ociosidade do espaço físico, de professores, de funcionários e de equipamentos, o que, nas instituições públicas, se reflete em desperdícios dos investimentos do governo e, nas particulares, redução no recebimento de mensalidade. Para os estudantes, por sua vez, a evasão pode representar o atraso ou cancelamento de um sonho, perda de oportunidades de trabalho, de crescimento pessoal e de melhoria de renda, entre muitas outras consequências.

Em [1], os autores afirmam que há múltiplos fatores que levam ao abandono escolar nas instituições de ensino, podendo estar relacionados aos contextos sociais, culturais, políticos e econômicos.

O Anuário Brasileiro da Educação Básica [3] estabelece que evasão configura-se quando o aluno, após ter sido matriculado em determinado ano letivo, não se matricula na escola no ano seguinte, independentemente de sua condição de rendimento escolar ter sido de aprovado ou de reprovado. Enquanto abandono ocorre quando um estudante não conclui o ano letivo, ou seja, deixa de frequentar a escola sem terminar uma determinada série ou ano escolar.

De acordo com o "Relatório de Desenvolvimento Humano", elaborado pelo programa das Nações Unidas [4], o Brasil tem a terceira

maior taxa de abandono escolar (24,3%) entre os 100 países com maior Índice de Desenvolvimento Humano (IDH), só atrás da Bósnia Herzegovina (26,8%) e das ilhas de São Cristovam e Névis, no Caribe (26,5%). Na América Latina, somente a Guatemala (35,2%) e Nicarágua (51,6%) têm taxas de evasão superiores. As estatísticas apresentadas no relatório são referentes ao ensino primário. Visto que a evasão escolar ainda possui números alarmantes e que traz grandes impactos, pesquisas são realizadas com objetivo de entender os fatores que influenciam a evasão, de discutir as consequências deste fenômeno e de definir estratégias que visem minimizar essas ocorrências.

Tais fatores podem ser observados a partir dos diversos dados acadêmicos, de perfil e comportamento dos alunos. Esses dados estão em diferentes instituições educacionais e são armazenados em diferentes formatos e estruturas, demandando, portanto, soluções capazes de lidar com essas diferenças, pois a evasão é um fenômeno que atinge as mais diversas instituições de ensino, independentemente de nível, turno ou categoria administrativa.

Representar os dados escolares segundo um modelo conceitual padronizado pode fornecer subsídios para consultas mais homogêneas e análises mais elaboradas desses dados, mesmo quando oriundos de fontes distintas. Por exemplo, um modelo conceitual, como uma ontologia, representando o domínio escolar, mais especificamente as informações relevantes à evasão, proveria essa estrutura comum, facilitando a compreensão e o tratamento de dados nesse contexto. Assim, a ontologia poderia atuar como interlíngua a respeito deste domínio, permitindo maior interoperabilidade semântica ao lidar com os dados e facilitando a posterior utilização e/ou desenvolvimento de ferramental especializado de análise dos dados. De acordo com [5], ontologias são artefatos de representação da informação muito úteis para integração de dados e para garantir interoperabilidade semântica das informações que ela representa.

Em vista disso, este trabalho propõe uma Ontologia de Evasão Escolar para que a análise dos dados desse domínio possa ser realizada sobre diferentes fontes de dados. A fim de prover uma maneira uniforme de se analisar dados de evasão escolar, será também apresentada uma abordagem técnica para trabalhar os dados e permitir a realização de consultas homogêneas aos dados com diferentes origens. Tal abordagem se baseia no uso de ontologias para prover um melhor entendimento dos conceitos do domínio, padronizar os *schemas* de dados e fornecer um modelo útil para a elaboração de consultas. Essa possibilidade de realização de consultas mais homogêneas, a partir dos conceitos da ontologia, favorece um tratamento similar dos dados e a aplicação e construção de recursos de análise mais especializados.

Será apresentado um estudo de caso sobre o domínio de evasão escolar, com dados do Censo do INEP e da Plataforma Nilo Peçanha. O propósito é demonstrar uma das possibilidades de ferramentas de análise que podem ser aplicadas com o uso da abordagem proposta, a fim de melhorar a análise dos dados de evasão, para possibilitar uma melhor compreensão dos fatores, auxiliar na tomada de decisões e de medidas preventivas.

O artigo está organizado da seguinte forma: a Seção 2 descreve as principais noções do domínio de evasão escolar, tais como conceituações, causas, impactos e estratégias adotadas para minimizar esse fenômeno, bem como aspectos técnicos para definição de ontologias; a Seção 3 propõe uma abordagem para realização de análise de dados de diferentes fontes; a Seção 4 descreve um estudo de caso de como esses repositórios podem ser utilizados para análise de dados e descoberta de conhecimento sobre evasão escolar; a Seção 5 discute os trabalhos correlatos; e a Seção 6, as considerações finais.

2 EVASÃO ESCOLAR E ASPECTOS TÉCNICOS PARA DEFINIÇÃO DE UMA ONTOLOGIA DE DOMÍNIO

A literatura apresenta diversos conceitos de evasão escolar. Segundo [6], a evasão escolar é um fenômeno social, definido como a interrupção no ciclo de estudos. Já em [7], os autores apontam que a evasão escolar tem sido associada a situações tão diversas quanto a retenção e repetência do aluno na escola, a saída do aluno da instituição, a saída do aluno do sistema de ensino ou a não conclusão de um determinado nível de ensino. Para [8], a evasão é um fenômeno caracterizado pelo abandono do curso, rompendo com o vínculo jurídico estabelecido, não renovando o compromisso ou sua manifestação de continuar no estabelecimento de ensino. Em [9], é apresentado o resultado de um estudo realizado por uma Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras, designada pelo MEC. Nele, a evasão é definida como a saída definitiva do aluno de seu curso de origem, sem concluí-lo. Percebem-se diferentes amplitudes nos conceitos de evasão, mas de forma geral, tem-se entre eles a característica de uma interrupção no ciclo de estudos. Para o desenvolvimento da proposta deste trabalho, será utilizada uma conceituação baseada nas definições de [6] e [9], em que se tem a evasão escolar como um fenômeno social, definido como a interrupção no ciclo de estudos que dê origem à saída definitiva do sistema escolar.

De acordo com [10], o problema da evasão não é recente e está presente em todos os contextos educacionais, em maior ou menor grau. Por esse motivo, os estudos acerca do tema vêm se intensificando desde 1995, quando foi fundada a Comissão Especial de Estudos sobre Evasão, pelo Ministério da Educação. O interesse do Governo Federal em políticas de combate à evasão ocorre devido ao grande desperdício de recursos que esse problema gera, incluindo professores, funcionários, equipamentos e espaço físico, entre outros.

Diversos fatores podem levar à evasão escolar, tais como qualidade do ensino, ambiente escolar, fatores da vida social, questões familiares, distância da escola, falta de referencial familiar, exclusão social, bullying, horário de trabalho incompatível, desemprego, problemas financeiros, dentre outros. Em [11], são apresentados alguns motivos de evasão e de abandono, diretamente relacionados

a problemas na vida social do aluno, como a gravidez na adolescência, a necessidade de trabalhar para auxiliar na renda familiar ou a idade avançada decorrente de reprovações (desmotivação do aluno).

Já em relação às consequências da evasão escolar, sabe-se que são principalmente de cunho social e que refletem diretamente as desigualdades sociais. Em [12], é apresentada uma revisão da literatura, resumida na Tabela 1, em que são identificadas estratégias com objetivo de minimizar a evasão escolar. A tabela apresenta as estratégias identificadas e os artigos de referência. Percebe-se que, nas pesquisas que vêm sendo realizadas com objetivos de minimizar a evasão escolar, já foram propostas diversas estratégias pedagógicas com este objetivo. Acredita-se que, se essas estratégias forem aplicadas de forma direcionadas aos alunos com maiores riscos de evasão e a partir de uma análise dos reais fatores de evasão, as instituições podem alcançar melhorias significativas nos seus índices.

Tabela 1: Síntese das estratégias adotadas para minimizar a evasão. Fonte [12]

Estratégias	Fontes
- Suporte financeiro.	Diogo et al. (2016)
- Curso de nivelamento/Apoio pedagógico.	
- Oferta de eventos informativos.	
- Oferta de atividades extracurriculares, tais como: visitas técnicas, empresa júnior, eventos de ensino, pesquisa e extensão.	
- Garantir espaços de discussão sobre carreira, mercado de trabalho e opções de profissionalização.	
- Incentivo da parte docente e da coordenação aos alunos para participarem de programas de apoio pedagógico e de atividades extracurriculares; Orientação sobre a grade curricular antes da efetivação da matrícula.	Dias; Theóphilo; Lopes (2010)
- Orientação vocacional.	Kruger Júnior et al. (2011); Diogo et al. (2016)
- Oferta de monitorias/atendimento extraclasse pelos professores.	Tinto (1975/93)
- Realização de atividades integrativas em busca de fortalecer o vínculo aluno-professor. Ex.: festas, jantares comemorativos, encontro de confraternizações.	Kruger Júnior et al. (2011); Diogo et al. (2016)
- Acompanhamento e Intervenção da coordenação nos problemas de ensino-aprendizado desde as fases iniciais.	Andriola (2009); Dias; Theóphilo; Lopes (2010)
- Formação continuada do corpo docente.	Andriola (2009); Dias; Theóphilo; Lopes (2010)
- Oferta de infraestrutura adequada, tanto física (laboratórios, salas de aula, equipamentos, biblioteca etc.), como humana (valorização dos professores, remuneração para o desempenho das atividades de coordenação, auxílio de monitores/estagiários em salas com maior número de alunos etc.).	Andriola (2009); Dias; Theóphilo; Lopes (2010)
- Estrutura curricular flexível.	Andriola (2009); Dias; Theóphilo; Lopes (2010)
- Reforma da grade curricular: ampliação da carga horária de disciplinas com maior índice de reprovação; não oferta de disciplinas consideradas pesadas ou complexas no mesmo período; disponibilização das disciplinas em todos os semestres.	Andriola (2009); Dias; Theóphilo; Lopes (2010); Kruger Júnior et al. (2011); Diogo et al. (2016)
- Oferta de atividades de integração dos discentes (Ex.: recepção aos ingressantes; visitas de coordenadores às salas de aula etc.).	Tinto (1975/93)
- Estímulo para que haja integração à universidade e à comunidade acadêmica.	Tinto (1975/93); Bueno (1993); Silva Filho et al. (2007); Dias; Theóphilo; Lopes (2010)

Para obter um maior entendimento e embasamento a respeito desses fatores é importante representar o domínio de evasão escolar de maneira consistente, tanto para compreensão e comunicação dos aspectos de evasão, como para servir de base para análises mais aprofundadas. Com esta representação, por exemplo, pode-se identificar que a evasão em determinada Instituição Educacional tem maiores índices em cursos de determinada Área de Conhecimento, ou ainda, que é maior entre alunos do Gênero feminino no Turno noturno ou para alunos de determinada Faixa de Renda.

A fim de descrever o domínio e ter uma representação formal, são utilizadas ontologias que, segundo [13], são uma especificação formal e explícita de uma conceituação compartilhada. Baseado em [14], tem-se que ontologia é uma teoria lógica utilizada para capturar os modelos pretendidos de uma conceituação e excluir os

não pretendidos. Ou seja, uma teoria utilizada para especificar e explicitar uma conceituação. Nessa definição, percebe-se que uma ontologia está diretamente relacionada a uma representação de conceitos e suas relações em um determinado universo de discurso, que possibilita extrair os modelos conceituais mais adequados de um determinado domínio. Além disso, alguns recursos de Engenharia de Ontologias podem apoiar esta tarefa, tanto na definição de uma ontologia quanto na sua transformação em um artefato a ser processado computacionalmente.

Um desses recursos é UFO (*Unified Foundational Ontology*) [15], uma ontologia de fundamentação baseada em teorias de Ontologia Formal, Lógica Filosófica, Filosofia da Linguagem, Linguística e Psicologia Cognitiva, e desenvolvida com o intuito de prover uma fundamentação ontológica para linguagens gerais de modelagem conceitual. Uma vez que UFO classifica elementos de modelagem conceitual, ela permite representar conceitos e suas relações, e fornece embasamento para a criação de entidades chamadas *endurants*, objetos ordinários do cotidiano que persistem no tempo, por exemplo: *Pessoa, Aluno, Curso, Matrícula*. Segundo [16], ontologias de fundamentação são sistemas de categorias filosoficamente bem fundamentados e independentes de domínio que têm sido utilizados com sucesso para melhorar a qualidade de linguagens de modelagem e modelos conceituais. A ontologia proposta neste trabalho está fundamentada em UFO, a fim de se obter um modelo conceitual mais consistente e de maior qualidade.

Outro importante recurso utilizado foi *OntoUML (Ontological Unified Modelling Language)* [17], concebida como uma versão ontologicamente bem fundamentada de uma porção do diagrama de classes da *UML (Unified Modeling Language)* [18]. A ideia é empregar um método de engenharia de linguagem baseado em ontologia, para projetar uma linguagem para modelagem conceitual estrutural (ou seja, para modelar aspectos relacionados a objetos, suas propriedades e suas relações). É uma linguagem para construção de modelos conceituais que são fundamentados em UFO. Segundo [19], *OntoUML* é composta de um conjunto de estereótipos que representam distinções ontológicas definidas em *UFO*. Dentre tais estereótipos estão tipos («*kind*»), papéis («*role*») e relações («*relator*»), dentre outros. Esta utilização ficará evidente na formalização da ontologia de domínio, na Seção 3.

3 UMA ABORDAGEM PARA ANÁLISE DE DADOS DE DIFERENTES FONTES

A abordagem proposta tem como peça fundamental uma Ontologia de Domínio. É a partir dela que os dados de diferentes fontes podem ser padronizados em repositórios usando o mesmo *schema*, permitindo consultas e análises mais homogêneas. A Figura 1 mostra as macro etapas desta abordagem.

- (1) Modelar a Ontologia - dá origem a uma Ontologia de Referência, construída a partir da conceituação do domínio.
- (2) Obter Fontes de Dados - consiste em acessar / capturar os dados disponíveis do domínio.
- (3) Mapear *Schemas* - realiza o mapeamento semântico entre a Fonte de Dados (2) e a Ontologia de Referência (1), indicando qual é a relação dos tipos dos dados com os correspondentes conceitos e propriedades da ontologia.

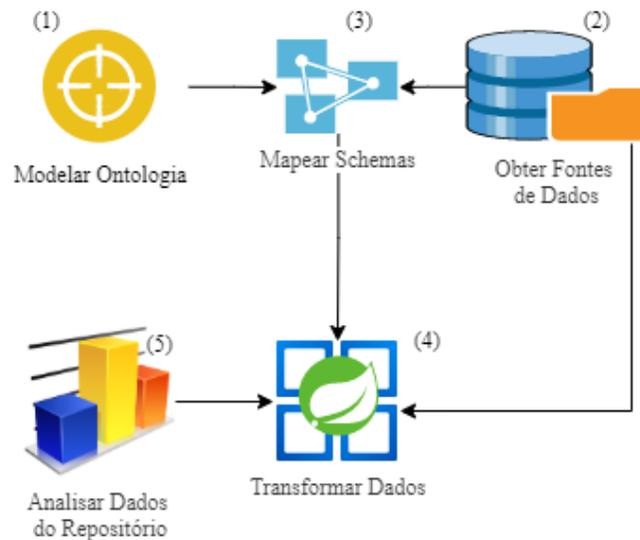


Figura 1: Abordagem proposta para realizar análise de dados de diferentes fontes

- (4) Transformar Dados - executa o processamento da padronização do *schema* de dados, o que dá origem ao repositório populado com instâncias da ontologia. Os dados variam conforme sua fonte, mas o formato (*schema*) segue sempre a ontologia.
- (5) Analisar Dados do Repositório - utilização de ferramentas que permitam a realização de análises de dados do domínio.

Com a utilização da ontologia como interlíngua, os mesmos tipos de informações podem ser extraídos de forma transparente a partir de diferentes fontes de dados. Por exemplo, aplicando a abordagem proposta para uma Instituição A, será gerado um repositório padronizado pela ontologia para essa instituição; de posse desse repositório podem ser feitas consultas para extrações de informações acerca de seus dados. Em seguida, pode-se aplicar a abordagem para uma Instituição B e com isso será gerado o seu próprio repositório; como o formato do repositório é padronizado pela ontologia, as mesmas consultas podem ser utilizadas também para análise de dados da Instituição B.

Dessa forma, a Ontologia desenvolvida apoia tanto a padronização dos dados, armazenados sob um mesmo *schema*; quanto as consultas a serem realizadas que, por utilizarem a mesma conceituação, podem ser feitas sobre qualquer dos repositórios gerados. Essa uniformidade das consultas e o modelo comum provido pela ontologia, favorecem o desenvolvimento e aplicação de ferramentas de análise de dados mais especializadas no domínio, independentemente da origem e formato inicial dos dados.

3.1 Ontologia do Domínio de Evasão Escolar

O primeiro artefato necessário na aplicação da abordagem é a ontologia (passo 1), que foi construída segundo o método SABiO [20]. A Ontologia de Evasão Escolar representa os principais conceitos e suas relações no domínio. Ela será utilizada para abstrair a origem

e formato dos dados e para gerar repositórios padronizados que permitam a aplicação de análises de dados para diferentes fontes.

Foram definidos os requisitos a fim de elaborar e validar a ontologia, escritos na forma de questões de competência (CQs) [21]. Alguns exemplos de CQs são:

- CQ01 - Quais alunos evadiram de uma instituição?
- CQ02 - Qual é a área de conhecimento de um curso?
- CQ03 - Quais as modalidades de ensino ofertadas por uma instituição?
- CQ04 - Qual o gênero de um aluno?
- CQ05 - Qual a cor/raça de um aluno?
- CQ06 - Quais os índices de evasão de uma instituição por faixa de renda?
- CQ07 - Quais os índices de evasão por gênero em cada faixa de renda?

É fundamental que se tenha informações de quais alunos evadiram, pois é este fenômeno que se pretende entender com as análises de dados a serem realizadas, e esse requisito foi representado pela CQ01. Sabe-se ainda que algumas áreas de conhecimento e modalidades de ensino possuem índices de evasão maiores que outras, endereçados pelas CQ02 e CQ03. As CQ04 e CQ05 estão relacionadas a características pessoais, o que pode contribuir para identificação de padrões de evasão. Já as CQ06 e CQ07 permitem visualizar números de evasão, relacionados a dois fatores, gênero e renda.

As CQs foram definidas de forma a cobrir uma série de características que podem ser úteis para a descoberta de padrões relacionados a evasão escolar, bem como questões que possibilitem validar a consistência da ontologia e, por conseguinte, dos dados. Além disso, são úteis na posterior definição de consultas sobre o repositório e de requisitos para ferramentas de análise.

Seguindo as boas práticas da engenharia de ontologias, foi realizada uma modularização da ontologia que está representada na Figura 2.



Figura 2: Modularização da ontologia

O módulo Organização (cor amarela) é responsável por representar conceitos e relações que dizem respeito às diferentes instituições de ensino. Por exemplo, o nível escolar da instituição, se é da educação básica ou superior, e sua categoria administrativa, se é pública ou privada. O módulo Socioeconômico (cor azul) representa os dados socioeconômicos dos alunos, por exemplo, gênero, cor/raça e renda. E o módulo Acadêmico (cor verde) provê distinções a respeito dos dados acadêmicos, por exemplo, curso, modalidade de ensino e matrícula. Foi utilizada UFO-A [17] para concepção e formalização da ontologia e os conceitos foram definidos a partir de referências como: Lei 9.394/1996 - Lei de Diretrizes e Bases da Educação; Decreto 9.235/2017 - que dispõe sobre o exercício das funções de regulação, supervisão e avaliação das instituições de educação superior; Constituição Federal Brasileira de 1988; glossário do Censo

da Educação Superior, realizado pelo INEP [22]; apresentação de resultados do Censo da Educação Superior [23]; tabela de classificação de áreas de conhecimento da Capes [24]; e publicações, tais como [25] e [26].

A Figura 3 apresenta uma versão parcial da Ontologia de Evasão Escolar (representada em OntoUML [17]). Tal versão descreve um recorte da ontologia mais ampla, ainda em desenvolvimento, voltado a análise dos aspectos mais relevantes para este artigo: fatores sócio econômicos básicos, características dos cursos e das matrículas.

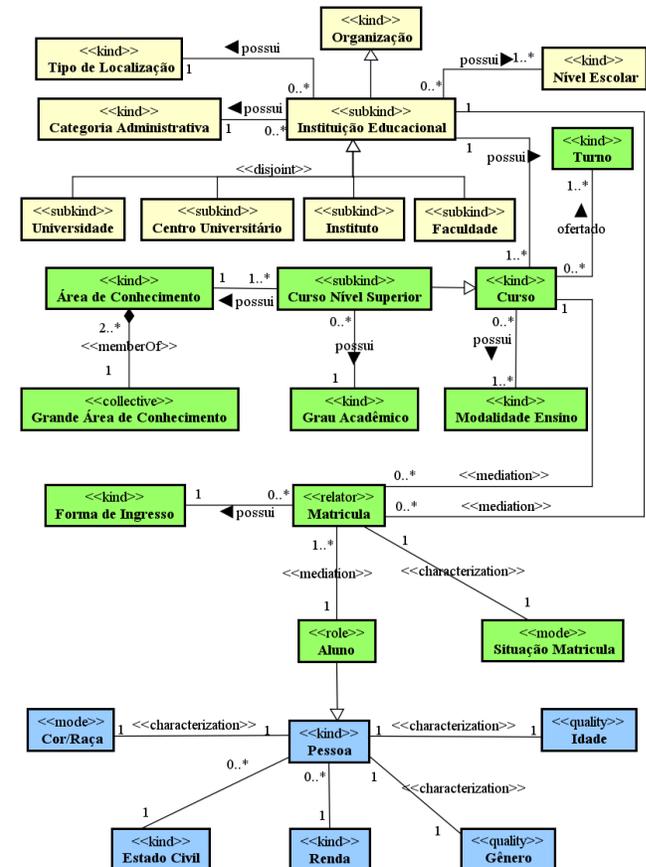


Figura 3: Ontologia de Evasão Escolar

No módulo Organização (em amarelo), tem-se o conceito de **Instituição Educacional**, que representa os tipos de organizações que estão inseridas no domínio, que são organizações que oferecem atividades educacionais. Neste estudo, o foco está no nível de ensino superior, sendo representado na ontologia por alguns subtipos de Instituição Educacional para tal nível de ensino: **Universidade**, **Centro Universitário**, **Instituto** e **Faculdade**. Por sua vez, essas instituições possuem uma **Categoria Administrativa** (pública federal, pública estadual, privada sem fins lucrativos, privada com fins lucrativos, dentre outras), um **Tipo de Localização** (urbana ou rural) e um ou mais **Níveis Escolares** (e.g., educação básica e educação superior). Para ilustrar as instâncias desses conceitos: o

Instituto Federal do Espírito Santo é uma instância de Instituto, que por sua vez é uma Instituição Educacional e uma Organização, de onde vem o princípio de identidade desta instância. Essa instituição possui Categoria Administrativa *pública federal* e um dos seus Níveis de Ensino é a *educação superior*.

No módulo Socioeconômico (em azul), o principal conceito é de **Pessoa**, e aparecem algumas características das pessoas que podem ser úteis para a identificação de padrões de evasão escolar: **Cor/Raça, Estado Civil, Idade, Gênero e Renda**. Por exemplo, *João da Silva* é uma instância de Pessoa, que possui Estado Civil *solteiro*, *Cor parda*, Renda de *R\$ 1500*, Idade de *23 anos* e Gênero *masculino*.

No módulo Acadêmico (em verde), os principais conceitos são **Curso, Aluno e Matrícula**. O Curso possui **Modalidade de Ensino** (presencial, a distância) e pode ser ofertado em diferentes **Turnos** (integral, matutino, vespertino), possui um **Grau Acadêmico** (Licenciatura, Bacharelado) e uma **Área de Conhecimento** (Engenharia Civil, Medicina) que é membro de uma **Grande Área de Conhecimento** (Engenharias, Ciências da Saúde).

O **Aluno** é o papel (*role*) que uma pessoa assume ao se matricular em um curso de uma instituição de ensino. A **Matrícula**, é responsável por estabelecer uma relação (*relator*) de vínculo entre o **Aluno**, o **Curso** e a **Instituição Educacional**, ou seja, permite representar que uma **Pessoa** se tornou **Aluno** de um **Curso** em uma determinada **Instituição**. Além disso, **Matrícula** possui uma **Forma de Ingresso** (Vestibular próprio, Prouni, Sisu) e tem uma característica que é a **Situação da Matrícula** (Desvinculado do curso, Formado, Cursando), por meio da qual se saberá, por exemplo, se um aluno evadiu ou concluiu.

Como exemplos de instâncias desses conceitos tem-se o Curso de Nível Superior *Bacharelado em Sistemas de Informação*, que possui a Modalidade de Ensino *presencial*, ofertado no Turno *vespertino*, o seu Grau Acadêmico é *Bacharelado*, possui área de conhecimento *Ciência da Computação*, que é membro da Grande Área de Conhecimento *Ciências Exatas e da Terra*. *João da Silva* é aluno desse curso e possui Matrícula *20211BSI0005*, que está com a Situação de Matrícula *Cursando*, tendo ingressado na instituição por meio do *Sisu*.

Foi definida uma página com as definições de alguns conceitos¹ da Ontologia de Evasão Escolar e as suas respectivas referências.

3.2 Repositórios de instâncias da ontologia

Como Fontes de Dados para execução da abordagem (passo 2), foram utilizados microdados do ano base 2019, da Plataforma Nilo Peçanha (PNP)² e do Censo da Educação Superior do INEP³. Essas bases foram selecionadas porque estão disponibilizadas como dados abertos, em bom nível de detalhes e possuem dados de diversas instituições de ensino, o que possibilita um maior alcance no que diz respeito a análise de dados de evasão escolar.

Foi definido um padrão de mapeamento (passo 3) dos dados para os conceitos da ontologia. Inicialmente, a estrutura dos dados de

cada fonte foi estudada para melhor compreensão. Em seguida, em cada base de dados (PNP e Censo), para cada conceito, foi identificado o dado com semântica correspondente, conforme ilustra a Figura 4. Também foram adicionados os índices dos arquivos originais para que o algoritmo de transformação dos dados possa gerar os repositórios (.CSV) já mapeados conforme os conceitos da ontologia.

Conceito da Ontologia	Índice Censo	Nom e Coluna Censo	Índice PNP	Nom e Coluna PNP
Genero	3	TP_SEXO	0	Sexo
Renda	-	-	1	Renda Familiar
CorRaça	4	TP_COR_RACA	2	Cor / Raça
Idade	2	NU_IDADE	3	Faixa Etária
Aluno	1	CO_ALUNO_CURSO	5	Código Matrícula
Matricula	17	DT_INGRESSO_CURSO	6	Data Ocorrencia Matrícula
SituacaoMatricula	11	TP_SITUACAO	7	Situação Matrícula
Turno	12	TP_TURNO	10	Turno
GrandeAreaConhecimento	13	NO_CINE_AREA_GERAL	19	Eixo Tecnológico
AreaConhecimento	14	NO_CINE_AREA_DETALHADA	20	Subeixo Tecnológico
ModalidadeEnsino	10	TP_MODALIDADE_ENSINO	21	Modalidade Ensino
Curso	5	NO_CURSO	24	Nome Curso
InstituicaoEducacional	6	NO_IES	30	Instituição
GrauAcademico	9	TP_GRAU_ACADEMICO	-	-

Figura 4: Exemplo de arquivo de mapeamento das bases para os conceitos da ontologia

Para geração dos repositórios, foi desenvolvida uma aplicação que transforma os dados (passo 4) em instâncias da ontologia a partir do mapeamento feito. As duas bases utilizadas para o experimento possuem grandes volumes de dados, reunindo informações de diversas instituições de ensino do Brasil, por isso foi importante a definição de uma arquitetura de alto desempenho para a criação dos repositórios.

Foi utilizado o *framework Spring Batch*⁴, aplicando uma técnica de processamento assíncrono em *threads* separadas para aumentar o desempenho. Neste contexto, foi criado um *job* que transforma os dados de entrada em instâncias da ontologia no repositório. Os dados de origem são lidos de um arquivo .CSV (*comma separated values*), em que cada linha é um item a ser processado. Esses itens são processados por várias *threads*, dando origem um novo arquivo .CSV que representa o repositório de instâncias da ontologia. A utilização do *Spring batch* favorece um desacoplamento das etapas de leitura, processamento e escrita. Com isso, se houver a necessidade de gerar os repositórios em outros formatos (e.g., *RDF - Resource Description Framework*, persistir em banco de dados, etc.), isso poderia ser feito de forma facilitada.

Esta arquitetura permitiu bons resultados no que diz respeito a desempenho. Para a base da PNP com 954 mil registros, o repositório foi criado em 1 minuto e 7 segundos. Já para a base do Censo da Educação Superior com 12,3 milhões de registros, com 32 *threads*, foi finalizado em 12 minutos e 57 segundos.

Para cada fonte de dados é gerado um repositório, todos seguindo o mesmo *schema*. A padronização dos *schemas* baseada na ontologia garante que os dados estejam representados segundo o mesmo formato, estrutura e semântica, permitindo que as análises sejam realizadas da mesma forma, mesmo para diferentes origens de dados.

¹Conceitos da Ontologia de Evasão Escolar: <https://github.com/ontologia/conceitos-evacao-escolar/wiki/Conceitos-da-Ontologia-Evas%C3%A3o-Escolar>

²Microdados PNP ano base 2019: <http://dadosabertos.mec.gov.br/pnp/item/118-2019-microdados-matriculas>

³Censo Educação Superior ano base 2019: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>

⁴Spring Batch: <https://docs.spring.io/spring-batch/docs/current/reference/html/index.html>

4 GERAÇÃO DE RELATÓRIOS PARA GESTORES ESCOLARES

Neste trabalho, a título de ilustração, foi realizado um experimento criando relatórios que demonstrem como as análises de dados (passo 5) podem ser realizadas para diferentes origens. Para isto, foi utilizado o Metabase⁵, uma ferramenta de *business intelligence* de código aberto. Ela permite realizar consultas sobre os dados, e exibe as respostas em formatos como um gráfico ou uma tabela detalhada. Para trabalhar com arquivos .CSV, foi utilizado o plugin Markeson⁶ que adiciona esta funcionalidade ao Metabase. Foi criado um painel no Metabase com algumas informações que são possíveis de obter a partir dos repositórios populados.

As Figuras 5 e 6 apresentam informações extraídas com uma consulta idêntica nos dois repositórios (PNP e Censo). Os gráficos ilustram como as análises podem ser realizadas. Apresentam os percentuais de alunos por situação de matrícula. Observa-se que para a base da PNP de 2019, com dados no âmbito da Rede Federal de Educação Profissional, Científica e Tecnológica, o percentual de evasão foi de 18,59% (desvinculado do curso) para essas instituições (figura 5). Já para a base do Inep, dados do Censo da Educação Superior de 2019, o percentual de evasão foi de 18,65% (desvinculado do curso, figura 6).

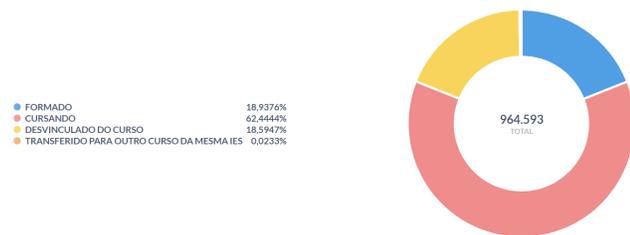


Figura 5: Análise de dados de matrículas base PNP 2019

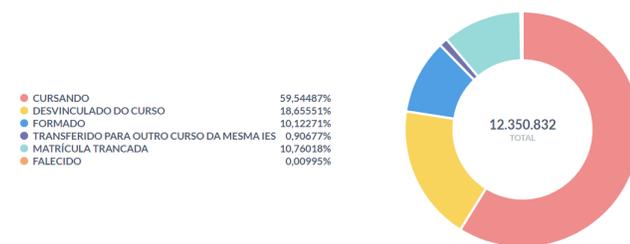


Figura 6: Análise de dados de matrículas base Censo 2019

Esses dois primeiros resultados extraídos dos repositórios foram elaborados para obter uma visão dos índices de evasão no ano de 2019 para as instituições que constam nas fontes de dados. Aqui nota-se que embora sejam dados de instituições diferentes, os percentuais de evasão ficaram bem próximos.

⁵Metabase: <https://www.metabase.com/start/>

⁶Plugin Markeson: <https://github.com/Markenson/csv-metabase-driver>

As duas bases selecionadas possuem informações de várias instituições brasileiras, com isso as análises aqui apresentadas podem ser aplicadas para diferentes instituições de ensino. Foi adicionada a possibilidade de filtrar os dados por instituição no painel, com isso os resultados apresentados a seguir (figuras 7, 8 e 10) são de um instituto (Instituto Federal do Espírito Santo) e de uma universidade (Universidade Federal do Espírito Santo), selecionados nas bases da PNP e do Censo respectivamente.

A Figura 7, mostra informações extraídas da base Censo com o objetivo de visualizar as quantidades de matrículas por situação e gênero. Observa-se que apesar do maior número de alunos do sexo feminino (14.294) em relação ao masculino (12.794), o número de alunos que evadiram do sexo masculino (1.379, ou seja, 10,8%) é maior do que do sexo feminino (1.175, 8,2%). Tal dado corrobora o estudo apresentado em [27], que demonstra que os homens apresentam taxa de evasão maior que as mulheres na educação superior. De posse desses tipos de informações, gestores escolares podem ir direcionando as estratégias pedagógicas que visem minimizar a evasão escolar para os perfis de alunos que possuem maiores riscos de evadirem.

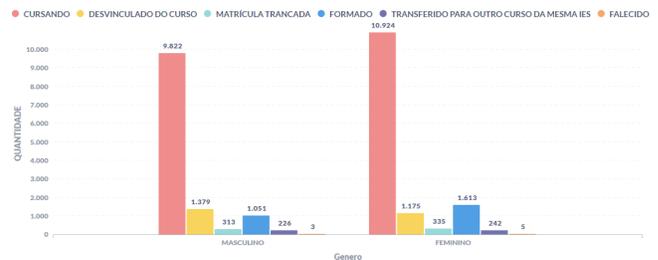


Figura 7: Situação de matrícula por gênero, instituição da base do Censo 2019

A Figura 8 mostra os percentuais de evasão por faixa de renda dos alunos do Instituto Federal do Espírito Santo, e permite concluir que os maiores índices de evasão ocorrem entre pessoas que possuem menores rendas. Em [28], também foram realizadas análise de dados de evasão escolar e foi possível observar que a modalidade com maior proporção de evadidos é a que abrange discentes oriundos de escola pública com renda de até 1,5 salário mínimo.

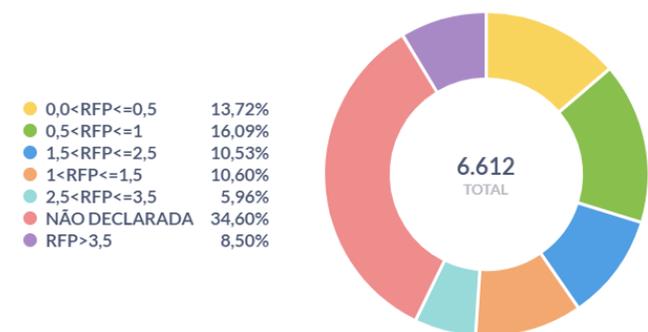


Figura 8: Dados de evasão por faixa de renda, instituição da base da PNP 2019

Uma última consulta, que pode ser visualizada na Figura 9, foi criada para extrair os dados de uma situação de matrícula, por gênero e renda, o que permite responder a CQ07 - Quais os índices de evasão por gênero em cada faixa de renda?. O resultado de sua execução com aplicação do filtro da situação "DESVINCULADO DO CURSO" é mostrado no gráfico da Figura 10.

```
select Renda, Genero, count(Aluno) as Quantidade
from dadosEvasao
[[where InstituicaoEducacional = {{InstituicaoEducacional}}
and SituacaoMatricula = {{SituacaoMatricula}}]]
group by Renda, Genero order by Renda
```

Figura 9: Exemplo de consulta

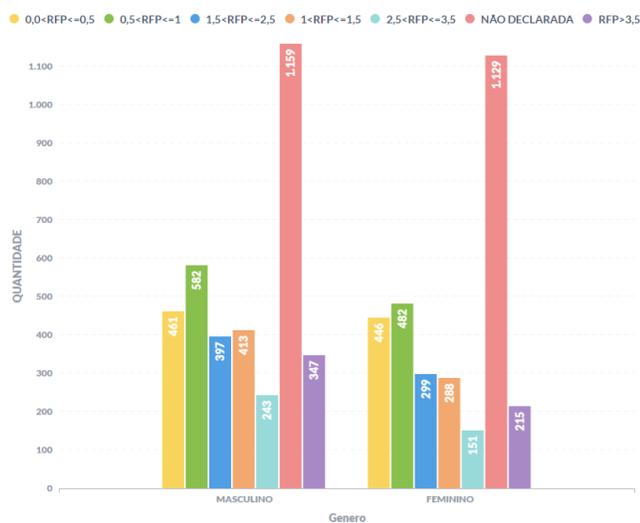


Figura 10: Dados de evasão por gênero e renda, instituição da base da PNP 2019

Os termos utilizados na consulta advêm da Ontologia de Evasão Escolar (**Genero**, **Renda**, **Aluno**, **SituacaoMatricula** e **InstituicaoEducacional**) independentemente da fonte de dados. Analisando o gráfico (Figura 10), percebe-se que, além da maior evasão nas menores faixas de renda, para todas as faixas de renda o número de evadidos é maior para alunos do sexo masculino.

Os dados apresentados são apenas exemplos dos tipos de informações que podem ser obtidas com o uso da abordagem proposta, no escopo selecionado. Com as mesmas consultas foi possível extrair informações para os dois repositórios criados a partir de origens de dados diferentes.

Vale destacar que embora a solução tecnológica adotada na demonstração tenha sido de repositórios CSV com análises a partir do Metabase, outras configurações também são possíveis na abordagem. A partir do mapeamento dos dados das fontes para a ontologia, podem ser gerados repositórios em RDF ou OWL para consultas sobre triplas, em um modelo relacional para consultas SQL, ou mesmo em formatos NoSQL. Tais opções permitem maior variabilidade na utilização e/ou desenvolvimento de ferramentas de análise.

5 TRABALHOS CORRELATOS

A literatura apresenta alguns trabalhos que, como este, buscam soluções para mitigação da evasão escolar. Nesse sentido, no trabalho [29], os autores apontam os principais fatores que influenciam o abandono escolar no ensino médio, e apresentam um diagrama de conceitos que divide esses fatores em três grupos de classificação: fatores ligados à escola, fatores relacionados ao aluno e fatores ligados à família. Nesse trabalho o objetivo do diagrama foi possibilitar ter um entendimento de como os conceitos estão relacionados e aprofundar o conhecimento a respeito dos motivos que levam alunos a abandonarem o ensino médio. A principal diferença em relação à nossa proposta é que aqui está sendo proposta uma ontologia como modelo conceitual, construída a partir das distinções fornecidas por uma ontologia de fundamentação (UFO). Isso permite, do ponto de vista de representação, um modelo mais fidedigno ao domínio de evasão escolar e a aplicação como interlíngua para análise de dados de diferentes fontes, e do ponto de vista técnico, maiores facilidades na preparação dos *schemas* e transformação de dados.

Outro trabalho correlato é o [5], no qual é apresentada a *Ontology For Learning Analytics (Onto4LA)*, criada com o objetivo de facilitar a integração de dados educacionais e garantir a interoperabilidade semântica. Embora haja correlação no que diz respeito ao uso de ontologias para permitir análise de dados de diferentes fontes escolares (a Onto4LA lida com conceitos como: *Interaction, Profile, Message, Log, etc.*), essa difere de nossa proposta por ter sido projetada para análise de dados da educação a distância, sendo modelada em uma perspectiva de eventos, enquanto a Ontologia de Evasão Escolar está mais focada em aspectos estruturais da educação em geral.

O trabalho desenvolvido em [30], embora não faça uso de ontologias para análise de dados, teve como objetivo de estudo identificar fatores que influenciam na decisão do aluno evadir ou concluir um curso de graduação em uma Universidade Federal. O artigo se baseou na teoria proposta em [31], que desenvolveu um modelo que serviu de inspiração e base para outros modelos teóricos sobre o assunto. A principal diferença para a proposta deste artigo está no fato de que aqui é possível a realização de análises mais amplas, que se dá pela definição de um modelo homogêneo e das transformações que foram propostas.

6 CONSIDERAÇÕES FINAIS

Sabe-se que os impactos da evasão escolar são grandes, para a sociedade, para as instituições de ensino e principalmente para os alunos. Segundo [32], as consequências da evasão escolar na vida dos educandos têm sido visíveis na sociedade contemporânea, trazendo altos índices de violência, criminalidade e envolvimento com drogas, comprometendo seus sonhos e seus projetos futuros. Com isso, é muito importante a criação de recursos tecnológicos capazes de contribuir para reduzir a evasão.

Acredita-se que análise de dados sobre a evasão escolar é algo fundamental para contribuir para um melhor entendimento dos fatores que levam à evasão, bem como para definição de políticas e estratégias que visem minimizar esse fenômeno. Essas análises permitem a gestores escolares definirem ações mais direcionadas aos perfis de alunos que mais evadem ou situações que apontem para maior risco de evasão. Por exemplo, de posse de repositórios

de dados como os propostos, os gestores podem visualizar que os maiores índices de evasão estão em cursos de determinada área de conhecimento, ou entre alunos de baixa renda, conforme visto no estudo de caso apresentado.

A definição de uma ontologia de referência bem fundamentada cria uma representação fidedigna e um vocabulário comum a respeito do domínio de Evasão Escolar, facilitando a comunicação, atribuindo maior expressividade semântica para os repositórios criados e permitindo que as análises sejam realizadas para diferentes fontes de dados.

Além da Ontologia de Evasão Escolar, este trabalho apresentou uma abordagem que envolve o mapeamento e transformação eficiente dos dados para uma base consistente e padronizada, de onde podem ser realizadas consultas homogêneas. Ferramentas adicionais de consulta ou análise podem ser aplicadas ou desenvolvidas com base na ontologia, independentemente da fonte de dados.

Em suma, ocorre a aplicação de informática na educação, a fim de contribuir para que os índices de evasão sejam minimizados na educação brasileira. Como perspectivas futuras, tem-se a ampliação da versão da Ontologia Evasão Escolar apresentada, visando acrescentar outros conceitos relevantes, por exemplo, instituições educacionais de outros níveis e dados acadêmicos de nota e frequência.

Também está sendo explorada a utilização de outras tecnologias a partir dos repositórios gerados que permitam formas mais elaboradas de análise, por exemplo, a aplicação de técnicas de *Machine Learning (ML)* sobre os repositórios. Assim será possível gerar dinamicamente modelos de ML para predição de diversos aspectos relacionados à evasão escolar. De fato, complementar a abordagem proposta neste trabalho para aplicação de ML é algo promissor, uma vez que as soluções que fazem uso dessas técnicas são geralmente aplicadas para uma base específica, enquanto a abordagem proposta é focada em permitir a análise para múltiplas fontes de dados.

REFERÊNCIAS

- [1] Cristiane Aparecida Dos Santos Baggi and Doraci Alves Lopes. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16(2):355–374, 2011.
- [2] Gérson Tontini and Silvana Anita Walter. Pode-se identificar a propensão e reduzir a evasão de alunos?: ações estratégicas e resultados táticos para instituições de ensino superior. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 19(1):89–110, 2014.
- [3] ANUÁRIO BRASILEIRO DA EDUCAÇÃO BÁSICA. Todos pela educação. São Paulo, 2019.
- [4] Raimundo Barbosa Silva Filho and Ronaldo Marcos de Lima Araújo. Evasão e abandono escolar na educação básica no brasil: fatores, causas e possíveis consequências. *Educação por escrito*, 8(1):35–48, 2017.
- [5] Luiz Carlos Carchedi, Jairo Souza, Eduardo Barrère, and Fabricio Mendonça. Onto4la: uma ontologia para integração de dados educacionais. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 7, page 439, 2018.
- [6] Natália Pacheco de Lacerda Gaioso. O fenômeno da evasão escolar na educação superior no brasil. *Brasília, DF: Universidade Católica de Brasília*, page 20, 2005.
- [7] Rosemary Dore and Ana Zuleima Lüscher. Permanência e evasão na educação técnica de nível médio em minas gerais. *Cadernos de pesquisa*, 41:770–789, 2011.
- [8] Elaine de Jesus Melo de Araújo et al. Evasão no proeja: estudo das causas no instituto federal de educação, ciência e tecnologia do maranhão/ifma. 2012.
- [9] MEC Brasil. Comissão especial de estudos sobre a evasão nas universidades públicas brasileiras. <http://www.dominiopublico.gov.br/download/texto/me001613.pdf>, 15(01):2007, 1997.
- [10] Mariana Andressa Luna Pinheiro, Josenildo Costa da Silva, and Bruno Feres de Souza. Aprendizado de máquina aplicado à análise de evasão no ensino superior. *Anais do Computer on the Beach*, pages 512–521, 2018.
- [11] Marcelo Neri et al. Motivos da evasão escolar. 2015.
- [12] Izaqueline Jhusmicle Alcântara da Silva, Gilberto José Miranda, Edvalda Araujo Leal, and Janser Moura Pereira. Estratégias das coordenações dos cursos de ciências contábeis para combater a evasão. *Revista Universo Contábil*, 14(2): 61–81, 2019.
- [13] Rudi Studer, V Richard Benjamins, and Dieter Fensel. Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2):161–197, 1998.
- [14] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In *Handbook on ontologies*, pages 1–17. Springer, 2009.
- [15] Veruska Carretta Zamborlini. Estudo de alternativas de mapeamento de ontologias da linguagem ontouml para owl: Abordagens para representação de informação temporal. *Federal University of Espírito Santo. Available only in Portuguese*, 2011.
- [16] G Guizzardi, RA Falbo, and RSS Guizzardi. A importância de ontologias de fundamentação para a engenharia de ontologias de domínio: o caso do domínio de processos de software. *IEEE Latin America Transactions*, 6(3):244–251, 2008.
- [17] Giancarlo Guizzardi. Ontological foundations for structural conceptual models. 2005.
- [18] Grady Booch. *UML: guia do usuário*. Elsevier Brasil, 2006.
- [19] Lucas Pereira, Rodrigo F Calhau, Paulo Sérgio dos Santos Júnior, and Mateus B Costa. Ontologia de domínio de doação de órgãos e tecidos para apoio a integração semântica de sistemas. In *CIbSE*, page 222, 2015.
- [20] Ricardo de Almeida Falbo. Sabio: Systematic approach for building ontologies. In *ONTO. COM/ODISE@ FOIS*, 2014.
- [21] Michael Grüninger and Mark S Fox. Methodology for the design and evaluation of ontologies. 1995.
- [22] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Glossário do censo da educação superior 2013. *Sinopse Estatística*, 2014.
- [23] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Apresentação dos resultados do censo da educação superior 2017. *Sinopse Estatística*, 2018.
- [24] DE APERFEIÇOAMENTO DE PESSOAL DE COORDENAÇÃO and DE PESSOAL-CAPEs. Tabela de áreas de conhecimento/avaliação. *Brasília, DF*, 2014.
- [25] Afrânio Mendes Catani and JF de OLIVEIRA. A educação superior. *Organização do ensino no Brasil: níveis e modalidades na Constituição Federal e na LDB*, 2:73–84, 2002.
- [26] Cássio de Almeida Lima, Maria Aparecida Vieira, Fernanda Marques da Costa, Jucimere Fagundes Durães Rocha, and Orlene Veloso Dias. Correlação entre perfil sociodemográfico e acadêmico e formas de ingresso na graduação em enfermagem. *Revista de Enfermagem UFPE on line*, 9(4):7986–7994, 2015.
- [27] Melina Klitzke. Desigualdades de resultados na educação superior: primeiras análises da evasão de curso na ufjf. *Desafios para o ensino superior brasileiro no contexto contemporâneo*, 22290:61, 2020.
- [28] Antonio Waneton Paulo Pinheiro Sousa and Miliana Augusta Pereira Sampaio. Principais elementos que contribuem para a evasão do ensino superior: Um estudo bibliográfico. *Humanidades & Inovação*, 7(17):74–86, 2020.
- [29] Tufi Machado Soares, Neimar da Silva Fernandes, Mariana Calife Nóbrega, and Alexandre C Nicolella. Fatores associados ao abandono escolar no ensino médio público de minas gerais. *Educação e Pesquisa*, 41(3):757–772, 2015.
- [30] Jaime Souza Sales, Gutemberg Hespanha Brasil, Teresa Cristina Janes Carneiro, and Maria Auxiliadora de Carvalho Corassa. Fatores associados à evasão e conclusão de cursos de graduação presenciais na ufes. *Meta: Avaliação*, 8(24): 488–514, 2016.
- [31] Vincent Tinto. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125, 1975.
- [32] Vivian Leamari Magalhães Bezerra. Narrativas das histórias de vida da evasão escolar dos estudantes da educação de jovens e adultos–eja. 2019.