

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DAINF - DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

MATEUS GUIMARÃES BELIZARIO

**LINKED OPEN DATA EM CIDADES INTELIGENTES: O
CASO DE MOBILIDADE E EDUCAÇÃO EM CURITIBA**

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA
2020

MATEUS GUIMARÃES BELIZARIO

**LINKED OPEN DATA EM CIDADES INTELIGENTES: O
CASO DE MOBILIDADE E EDUCAÇÃO EM CURITIBA**

Proposta de Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Bacharel.

Orientadora: Prof^a. Dra. Rita Cristina Galarraga Berardi
DAINF - Departamento Acadêmico de Informática -UTFPR

CURITIBA
2020

MATEUS GUIMARÃES BELIZARIO

**LINKED OPEN DATA EM CIDADES INTELIGENTES: O CASO DE MOBILIDADE E
EDUCAÇÃO EM CURITIBA**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título
de Bacharel em Sistemas de Informação da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Data de aprovação: 11/Novembro/2020

HELOISE MANICA PARIS TEIXEIRA

Doutorado

Universidade Tecnológica Federal do Paraná (UTFPR)

REGIANE DA SILVA MACUCH

Doutorado

Universidade Tecnológica Federal do Paraná (UTFPR)

RITA CRISTINA GALARRAGA BERARDI

Doutorado

Universidade Tecnológica Federal do Paraná (UTFPR)

CURITIBA

2020

A minha família por toda dedicação em fazer a educação sempre presente na minha vivência, aos meu amigos por serem meu suporte durante a jornada acadêmica e aos professores pelos ensinamentos, correções e apoio.

AGRADECIMENTOS

Agradeço a Universidade Tecnológica Federal do Paraná, especificamente ao departamento acadêmico de Informática por todo apoio e suporte durante os anos de graduação e na caminhada científica e acadêmica.

Uma Cidade Inteligente é uma cidade que dá inspiração, compartilha cultura, conhecimento e vida, uma cidade que motiva seus habitantes a criar e fluir em suas próprias vidas (RIOS, 2012).

RESUMO

BELIZARIO, Mateus Guimarães. Linked Open Data em Cidades Inteligentes: O Caso de Mobilidade e Educação em Curitiba. 2020. 118 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

A rápida transição para uma população altamente urbanizada faz com que governos enfrentem novos desafios sobre o gerenciamento de dados e informações em espaços urbanos. Um deles diz respeito à necessidade de modelos de informações de semelhante semântica e a capacidade de compartilhar e conectar informações de fontes e em formatos heterogêneos, permitindo a melhora nas tomadas de decisões operacionais por gestores e cidadãos. A Lei de Acesso à Informação (LAI) fez com que a política de abertura de dados governamentais recebesse maior atenção. Esses dados podem ajudar no aumento da transparência e a integridade do setor público, assim como permitir a inclusão e independência dos cidadãos sobre produção e utilização desses dados. Neste sentido, o objetivo deste trabalho é integrar dados abertos, utilizando ontologias, sobre os domínios de educação e mobilidade da cidade de Curitiba. Os dados obtidos são um indício da relação entre o desempenho acadêmico dos estudantes do município e as características de mobilidade urbana da cidade.

Palavras-chave: Ontologia. Dados Abertos. Dados Conectados. Cidades Inteligentes.

ABSTRACT

BELIZARIO, Mateus Guimarães. Linked Open Data in Smart Cities: Mobility and Education in Curitiba. 2020. 118 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

The fast transition to a highly urbanized population causes some governments to face new challenges in managing data and information in urban spaces. One of the concerns is the need for common information models and the ability to share and connect information from sources and in heterogeneous formats, allowing the improvement of operational decision making by managers and citizens. The Law on Access to Information (LAI) has made the policy of opening government data receive more attention. Such data can be helpful to increase the transparency and integrity of the public sector as well as enable citizens to be included and independent in their production and use. The objective of this paper is to integrate open data, using ontologies, on the domain of education and mobility of the city of Curitiba. The data obtained is an indication of the relationship between the academic performance of students in the municipality and the characteristics of urban mobility in the city.

Keywords: Ontology. Open Data. Linked Data. Smart Cities.

LISTA DE FIGURAS

Figura 1 – Artigos incluídos no corpo do trabalho em cada fase de seleção da revisão.	21
Figura 2 – Frequência de publicação dos artigos selecionados.	22
Figura 3 – Porcentagem dos artigos por categoria.	23
Figura 4 – Cenário descrevendo um conjunto de triplas e seus relacionamentos.	37
Figura 5 – Metodologia do estudo	43
Figura 6 – Transporte Utilizados por Estudantes	46
Figura 7 – Importância do Transporte Público	47
Figura 8 – Vulnerabilidades na Mobilidade Urbana - Estudantes	48
Figura 9 – Perfil de Especialistas de Educação	49
Figura 10 – Valor da Visão Integrada dos Domínios	50
Figura 11 – Vulnerabilidades na Mobilidade Urbana - Educação	51
Figura 12 – Nuvem de Palavras: Educação e Mobilidade Urbana	52
Figura 13 – Vocabulário Definido	56
Figura 14 – Dados abertos utilizados	57
Figura 15 – Mapa de Relações entre Classes	63
Figura 16 – Questão de Competência 1	71
Figura 17 – Resultado: Questão de Competência 1	72
Figura 18 – Questão de Competência 2	72
Figura 19 – Resultado: Questão de Competência 2	73
Figura 20 – Questão de Competência 3	73
Figura 21 – Resultado: Questão de Competência 3	74
Figura 22 – Questão de Competência 4	74
Figura 23 – Resultado: Questão de Competência 4	75
Figura 24 – Desempenho Médio no ENEM por Pontos de Acesso Próximos	75
Figura 25 – Máximo Desempenho no ENEM por Pontos de Acesso Próximos	76
Figura 26 – Consulta de Acessos por Bairro de Curitiba	77
Figura 27 – Quantidade de Acessos a Transporte Público por Bairro	77
Figura 28 – Consulta de Acessos por Tipo de Instituição	78

LISTA DE QUADROS

Quadro 1 – Lista de Classes	60
Quadro 2 – Lista de Classes	61
Quadro 3 – Lista de Data Properties	62

LISTA DE TABELAS

Tabela 1 – Artigos por Repositório	20
Tabela 2 – Definições de Cidades Inteligentes	34
Tabela 3 – 5 estrelas para Dados Abertos Conectados	41
Tabela 4 – Lista de Tabelas Relacionais	65

LISTA DE ABREVIATURAS E SIGLAS

AMCIS	<i>Americas Conference on Information Systems</i>
ARM	<i>Association Rule Mining</i>
BRT	<i>Bus Rapid Transit</i>
DAG	Dados Abertos Governamentais
DOT	Desenvolvimento Orientado ao Trânsito
DLs	<i>Description Logics</i>
DSL	<i>Domain Specific Language</i>
HTML	<i>HyperText Markup Language</i>
ICF	<i>Intelligent Community Forum</i>
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IoT	<i>Internet of Things</i>
INSE	Indicador de Nível Socioeconômico das Escolas
IPPUC	Instituto de Pesquisa e Planejamento Urbano de Curitiba
LAI	Lei de Acesso à Informação
OWL	<i>Ontology Web Language</i>
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes</i>
RDFS	<i>Resource Description Framework Schema</i>
RFID	<i>Radio-Frequency Identification</i>
TIC	Tecnologia de Informação e Comunicação
URI	<i>Universal Resource Identifier</i>
XML	<i>Extensible Markup Language</i>
W3C	<i>World Wide Web Consortium</i>

Lista de Algoritmos

1	Mapeamento de Endereços de Escolas	66
2	Mapeamento de Escolas	66
3	Mapeamento de Endereços de Universidades	67
4	Mapeamento de Universidades	67
5	Mapeamento ENADE	67
6	Mapeamento Prova Brasil	67
7	Mapeamento Prova ENEM	68
8	Mapeamento Estatísticas Educacionais	68
9	Mapeamento de Linhas de Ônibus	69
10	Mapeamento de Terminal de Ônibus	69
11	Mapeamento de Ponto de Ônibus	69
12	Mapeamento de Linha de Ônibus-Terminal de Ônibus	69
13	Mapeamento de Linha de Ônibus-Ponto de Ônibus	69
14	Mapeamento Distância Escola-Terminal/Ponto	70
15	Mapeamento Distância Universidade-Terminal/Ponto	70

SUMÁRIO

1 – INTRODUÇÃO	15
1.1 Problema e Objeto de Estudo	16
1.2 Objetivo Geral e Específicos	17
1.3 Organização do Trabalho	18
2 – ESTADO DA ARTE	19
2.1 Protocolo da Revisão Sistemática	19
2.1.1 Análise Geral Sobre os resultados da Revisão Sistemática	21
2.2 Resultados da Revisão Sistemática	23
2.2.1 Cidades Inteligentes: Problemas e Oportunidades	24
2.2.2 Engajamento Público	26
2.2.3 Soluções Aplicadas a Cidades Inteligentes	27
2.2.4 Discussão sobre os resultados da Revisão Sistemática	28
3 – TRABALHOS CORRELATOS	30
4 – FUNDAMENTAÇÃO TEÓRICA	33
4.1 Cidades Inteligentes	33
4.2 Mobilidade Urbana e Educação	35
4.3 <i>Linked Data</i>	36
4.4 Modelos Ontológicos	39
4.4.1 <i>Resource Description Framework</i>	39
4.5 Dados Abertos	40
5 – METODOLOGIA	43
5.1 Etapas de Desenvolvimento	43
6 – DESENVOLVIMENTO	45
6.1 Seleção de Especialistas de Domínio	45
6.2 Levantamento de Necessidades de Informação	45
6.2.1 Estudantes	46
6.2.2 Especialistas de Educação	48
6.2.3 Especialistas de Mobilidade Urbana	52
6.3 Questões de Competência	53
6.4 Definição de Vocabulário	54
6.5 Seleção de Ferramentas	56
6.6 Construção da Ontologia	58

6.6.1	Especificação da Ontologia	58
6.6.2	Aquisição de Conhecimento	59
6.6.3	Tratamento dos Dados	63
6.6.4	Conexão dos Dados dos Domínios	66
7	– RESULTADOS	71
8	– CONCLUSÃO	79
8.1	Trabalhos Futuros	80
8.2	Desdobramentos do Trabalho	80
	Referências	82
	 Apêndices	 88
	APÊNDICE A–Formulário: Mobilidade e Educação em Curitiba	89
	APÊNDICE B–Código: Tratamento de Dados Prova Brasil 2017	105
	APÊNDICE C–Código: Tratamento de Dados ENEM 2015	107
	APÊNDICE D–Código: Tratamentos de Dados ENADE 2017	108
	APÊNDICE E–Código: Tratamento de Relação Escolas-Terminais	110
	APÊNDICE F–Código: Tratamento de Relação Escolas-Pontos de Ônibus	112
	APÊNDICE G–Código: Tratamento de Relação Universidade-Terminal	114
	APÊNDICE H–Código: Tratamento de Relação Universidade-Ponto de Ônibus	116
	APÊNDICE I– Código: Tratamento de Pontos de Ônibus de Curitiba	118

1 INTRODUÇÃO

A rápida transição para uma população altamente urbanizada fez com que alguns governos enfrentassem novos desafios em relação a temas-chave como desenvolvimento sustentável, educação, energia, meio ambiente, segurança e serviços públicos, entre outros. A disponibilidade de Tecnologias de Informação e Comunicação (TICs) em cidades inteligentes estimula o desenvolvimento de novos serviços e aplicações, e cria um ambiente mais eficiente na solução colaborativa de problemas e inovação (BOLÍVAR, 2018).

Sobre a conceitualização do termo, Zanella et al. (2014) dizem que ainda não existe uma definição formal e amplamente aceita para cidades inteligentes, mas seu objetivo final é fazer um melhor uso dos recursos públicos, aumentando a qualidade dos serviços oferecidos aos cidadãos, enquanto reduz os custos operacionais das administrações públicas.

Segundo Lemos (2013), o conceito de cidades inteligentes surge neste contexto para estimular o ambiente público na tomada de decisões, aumentar os laços comunitários e a participação política. Apesar de todos os benefícios que a aplicação deste conceito pode proporcionar, existem dificuldades e dúvidas sobre como desenvolver e usar tecnologias para construir e manter cidades inteligentes (MARCEAU, 2008).

Em direção ao desenvolvimento de cidades inteligentes, Mellouli, Luna-Reyes e Zhang (2014) declaram que os governos ao longo dos anos se engajaram em um movimento para abrir dados com licenças abertas e em formatos mais fáceis de reutilizar. As soluções *smart data* podem ser usadas para permitir a capacidade de gerenciar estrategicamente os dados do governo por meio da sua abertura, distribuição e estruturação (ALGEMILI, 2016).

A aplicação de soluções sobre dados abertos pelas cidades pode ajudar na confiança dos serviços prestados, aumentando a transparência e a integridade do setor público. O uso desse conceito também pode possibilitar maior participação e inclusão da comunidade na tomada de decisões, além de permitir que os próprios cidadãos possam produzir e usar dados para fins públicos e privados. Assim como permitir que eles produzam suas próprias iniciativas com base nos dados abertos fornecidos para resolver ou aliviar problemas presentes nos diferentes domínios das cidades inteligentes (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

Nesse sentido, deve-se pensar no cidadão e sua independência como elemento central das cidades inteligentes. Sobre isso, Albino, Berardi e Dangelico (2015) dizem que uma cidade com bom desempenho na economia, nas pessoas, na governança, na mobilidade, no meio ambiente e na qualidade de vida, é construída com base na combinação de recursos e atividades de cidadãos independentes e conscientes.

Os principais desafios no gerenciamento de dados em cidades inteligentes, conforme Naphade et al. (2011), são a necessidade de modelos de informações comuns e a capacidade

de compartilhar informações de vários agentes e instituições detentoras desses dados dentro de uma cidade. Ainda segundo estes autores, para garantir a visibilidade ao gerenciar serviços e infraestruturas de cidades inteligentes, é necessário integrar dados de fontes diferentes, cada qual com sua própria frequência de amostragem, características, formatos e semântica.

Por exemplo, as informações relacionadas ao fluxo e mobilidade de cidadãos estão espalhadas por muitas instituições e domínios diferentes, incluindo transporte e planejamento urbano. Criar e aplicar um modelo unificado de informações possibilita obter um quadro mais completo da atividade urbana e sua situação. A capacidade de compreender essas combinações e relações pode permitir a melhora nas tomadas de decisões operacionais por gestores e cidadãos.

Assim, Chourabi et al. (2012) explicam que a disponibilidade de grandes quantidades de informações abre oportunidades para entender a atividade da cidade por meio de modelagem e análise de dados heterogêneos originados de diferentes domínios.

Para relacionar e compreender o valor desses dados, são necessários métodos e abordagens empregados pela tecnologia da informação e comunicação para lidar, analisar e interpretar grandes quantidades de dados de diferentes domínios e fontes. Como a criação de ferramentas que permitem a coleta e processamento de dados, colaboração baseada na web e atualização da inteligência coletiva dos cidadãos (SCHAFFERS et al., 2011).

As tecnologias da Web Semântica têm um potencial em fornecer a base para novos serviços eletrônicos, auxiliar na tomada de decisão e no desenvolvimento de novas soluções nos ecossistemas urbanos. O uso de ontologias, *Linked Open Data* (LOD) e outras tecnologias semânticas abrem novas possibilidades em cidades inteligentes, pois podem combinar informações de várias fontes para fins como estatísticas, análises, mapas e publicações, informar aos usuários quando as informações correspondem aos seus interesses e descrever produtos e serviços com mais precisão (BISCHOF et al., 2014).

Em visão da heterogeneidade de informações e conjuntos de dados dispostos por cidades inteligentes, é possível utilizar ontologias para a integração e conexão de dados. As ontologia surgem como uma forma de padronização e homogeneização dos dados levando em consideração sua semântica, em outras palavras, o significado dos dados dentro de seu domínio. Sendo, portanto, um modelo de dados utilizado para representar os conceitos, ou classes, de um domínio e seus relacionamentos (GUARINO; OBERLE; STAAB, 2009).

1.1 Problema e Objeto de Estudo

O problema direcionado neste trabalho é a falta de homogeneização sobre o formato dos dados e seus valores semânticos, a falta da relação existente entre os dados dos diferentes domínios das cidades inteligentes e as formas para interpretar essas relações de forma a compreender a situação do ambiente do qual foram originados.

Para isso, o objeto de estudo selecionado se refere à heterogeneidade dos dados

abertos da cidade de Curitiba, no estado do Paraná no Brasil. De forma específica, foram selecionados dois domínios da cidade para se trabalhar: Educação e Mobilidade. Esses domínios foram selecionados pois demonstram a dificuldade de gestão, heterogeneidade e integração dos dados dispostos pelas bases de dados abertas da cidade. E a integração deles podem permitir novas visões e desdobramentos sobre a situação da educação, em diferentes níveis, e da mobilidade na cidade de Curitiba.

Por exemplo, ambos os domínios apresentam informações sobre localização e essa informação pode ser relacionada entre eles. As instituições de ensino do município estão localizados em um determinado endereço, e as linhas de transporte público passam por certas ruas e os terminais de ônibus estão localizados em endereços específicos. Porém essas informações estão isoladas e definidas por termos diferentes. Logo, não estão conectados.

Além disso, segundo Heinlein e Shinn (2000), a influência da mobilidade estudantil na escolarização, no desenvolvimento cognitivo e no desempenho acadêmico tem sido estudada. Uma das relações demonstradas é de que quanto maior o esforço de deslocamento, menor são as conquistas e rendimentos do estudante durante sua vida acadêmica.

Estabelecer uma relação entre mobilidade e realização educacional é um problema complexo, no entanto há variáveis que podem ser consideradas. Por exemplo, estudos apontam que as crianças que se movem com frequência maior que a média para chegar até a escola são mais propensas a ser pobres, mais propensas a vir de um lar com um único pai, e são mais propensas a estar em um lar onde o morador está desempregado ou não conseguiu se formar no ensino médio (LONG, 1992).

1.2 Objetivo Geral e Específicos

Este trabalho propõe a construção de um modelo ontológico de dados para representar a relação entre os domínios de mobilidade e educação por meio de conceitos de *Linked Open Data*. Este modelo deve ser capaz de lidar com dados de diferentes domínios, em diferentes formatos e de diferentes origens. Além de fornecer informações de forma integrada de modo a auxiliar na gestão desses domínios. Com a finalidade de atingir o objetivo principal deste trabalho, alguns objetivos específicos foram definidos:

- I. Conhecer os dois domínios e identificar possíveis benefícios e pontos de integração.
- II. Explorar os dados com objetivo de realizar a representação dos dados de maneira integrada.
- III. Compreender como a tecnologia *Linked Open Data* pode ser implementada no contexto da cidade de Curitiba para os domínios de Educação e Mobilidade.
- IV. Compreender como *Linked Open Data* podem produzir novas informações relevantes para a cidade de Curitiba.

1.3 Organização do Trabalho

Esse trabalho está organizado de modo a apresentar no capítulo dois o estado da arte e o processo de revisão sistemática da literatura utilizado neste projeto para a seleção de trabalhos correlatos e construção da fundamentação teórica. O capítulo três apresenta os trabalhos correlatos encontrados na literatura que abordam questões sobre os problemas presentes em cidades inteligentes, o engajamento público nesses espaços e soluções aplicadas que partam de iniciativas *Linked Open Data*.

O capítulo quatro trata da fundamentação teórica, ela contém os conceitos importantes e relevantes para a compreensão deste trabalhos, suas propostas e discussões. Neste capítulo serão dispostos definições relacionadas a cidades inteligentes, *Linked Data*, *Open Data* e Modelos Ontológicos.

O capítulo cinco apresenta a metodologia do trabalho e os métodos utilizados para a organização e planejamento, assim como a descrição do processo de desenvolvimento .O capítulo seis explana os detalhes e processos do desenvolvimento do projeto. O capítulo sete contém os resultados obtidos e o capítulo oito a conclusão e os trabalhos futuros.

2 ESTADO DA ARTE

Este capítulo apresenta a revisão sistemática realizada por Belizario e Berardi (2019), publicada pela Conferência das Américas sobre Sistemas de Informação (AMCIS), para fundamentar este trabalho e seus processos.

A primeira seção deste capítulo explica o protocolo utilizado e a metodologia para a realização da revisão sistemática. A seção seguinte trata da análise geral sobre os resultados obtidos através da revisão sistemática e por fim apresentam-se as discussões oriundas deste processo.

Nesta seção, é descrito como foi realizado o processo de revisão sistemática deste trabalho, que foi realizada em janeiro de 2019. O principal objetivo desta revisão foi identificar soluções para a integração, interpretação e relação de dados gerados em espaços urbanos com o objetivo de incentivar o ambiente público na tomada de decisões, ampliando os laços comunitários, a participação política e a criação de novos conteúdos para a população.

2.1 Protocolo da Revisão Sistemática

Os repositórios Biblioteca Digital IEEE Xplore, Biblioteca Digital ACM, Scopus, Springer, Periódicos Capes e Google Acadêmico foram utilizadas para apoiar a revisão sistemática da literatura sobre como ocorre o uso de dados inteligentes em cidades inteligentes. Essa busca foi realizada em janeiro de 2019 e foi orientada pelo protocolo apresentado por Kitchenham (2004).

A pesquisa envolveu a expressão geral das palavras-chave ["*smart data*"] e ["*smart city*" ou "*smart cities*"], ou ["*smart data cities*"] e ["*open data*"], presentes em qualquer parte do trabalho.

Ao utilizar estas expressões, é possível ter acesso a artigos que utilizem as mais diversas técnicas relacionadas a este tema, gerando assim maior variedade para a pesquisa e, em seguida, verificando o estado da arte nessa área. Como um dos objetivos desta revisão foi identificar diferentes perspectivas em torno do uso de dados inteligentes, não foram utilizados quaisquer sinônimos possíveis do termo, como *big data* ou *linked data*. Na busca, usou-se apenas conceitos diretamente ligados a cidades inteligentes, dados inteligentes e dados abertos para entender como a literatura representa esses termos e quais são as ferramentas e abordagens utilizadas na construção de cidades suportadas por dados inteligentes.

Tabela 1 – Artigos por Repositório

Repositório	Artigos	String de Busca
IEEE Xplore Digital Library	1	((("smart data"AND ("smart city"OR "smart cities"))) OR "smart data cities") AND "open data"
ACM Digital Library	33	+(smart data) +("smart citysmart cities") +("open data")
Periodicos Capes	36	((("smart data"AND ("smart city"OR "smart cities"))) OR "smart data cities") AND "open data"
Google Acadêmico	355	((("smart data"AND ("smart city"OR "smart cities"))) OR "smart data cities") AND "open data"
Springer	36	((("smart data"AND ("smart city"OR "smart cities"))) OR "smart data cities") AND "open data"
Scopus	11	((("smart data"AND ("smart city"OR "smart cities"))) OR "smart data cities") AND "open data"

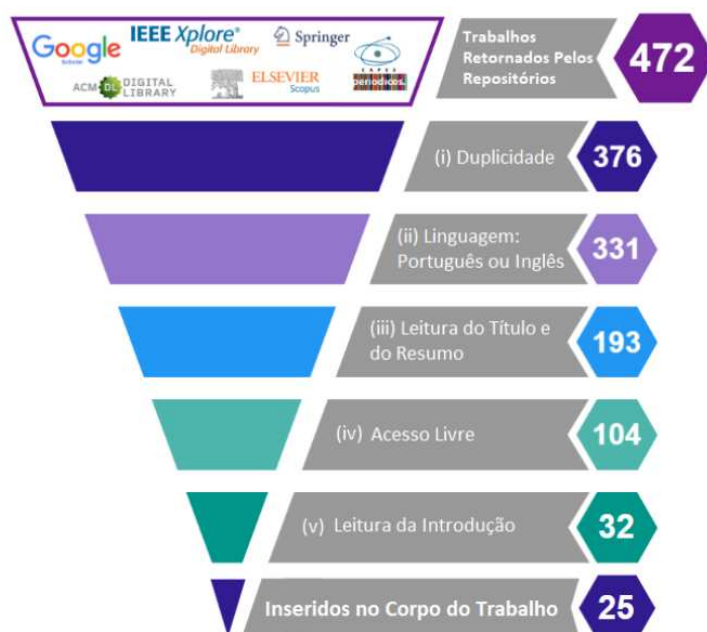
As publicações foram consideradas em inglês e português, entre os anos de 2014 e 2019. A língua inglesa foi selecionada pela abrangência e relevância da pesquisa e o português foi escolhido pela oportunidade de conhecer soluções e estudos locais. Este período foi estipulado porque os conceitos trabalhados *smart data* e cidades inteligentes são conceitos com diferentes significados para estudos anteriores à 2014, que estão muito mais relacionados majoritariamente à dispositivos *Internet of Things* (IoT). No total, foram devolvidos 472 artigos, conforme destacado na Tabela 1.

Na primeira fase, 96 artigos foram excluídos pelo primeiro critério, a duplicidade (i), eles estavam presentes em mais de uma biblioteca digital, o segundo (ii) critério é o idioma utilizado para escrever os artigos, somente foram selecionados trabalhos em português e inglês. E por este critério, foram excluídos 45 artigos. A leitura do título e resumo dos artigos foi o terceiro (iii) critério para a exclusão do trabalho. O título e o resumo deveriam apresentar discussões sobre soluções de dados abertos inteligentes para qualquer domínio de cidades inteligentes. Neste filtro foram descartados 138 artigos.

A partir dessa seleção preliminar, obteve-se 193 artigos restantes. Como quarto critério, (iv) o trabalho deve ser de livre acesso, o que fez com que 104 artigos fossem selecionados para a próxima fase de seleção e 89 artigos fossem descartados. E como último

critério (v), a introdução deveria estar relacionada às cidades inteligentes, envolvendo soluções para os problemas de informação e dados, usando dados abertos inteligentes. Neste filtro foram descartados outros 72 artigos. A Figura 1 ilustra o processo metodológico descrito.

Figura 1 – Artigos incluídos no corpo do trabalho em cada fase de seleção da revisão.



Fonte: Autoria própria

2.1.1 Análise Geral Sobre os resultados da Revisão Sistemática

Como resultado do processo de seleção e identificação dos artigos, vinte e cinco artigos preencheram os requisitos metodológicos descritos anteriormente. A Figura 2 mostra o ano de publicação desses artigos. Houve um crescimento constante de publicações sobre este tema desde 2014, atingindo o pico em 2018 com sete publicações, dentre os artigos selecionados.

Figura 2 – Frequência de publicação dos artigos selecionados.



Fonte: Autoria própria

Como forma de organizar a leitura e classificar os artigos selecionados, eles foram agrupados em três categorias diferentes percebidas durante a leitura dos artigos: Revisões teóricas, trabalhos sobre engajamento público e soluções para cidades inteligentes.

Na categoria de revisões teóricas foram incluídos onze artigos. Estes artigos descrevem conceitos relevantes sobre cidades inteligentes e seus desdobramentos, bem como sua implementação, problemas e dificuldades enfrentados. Além disso, apresentam tecnologias e ferramentas como propostas de soluções para alguns dos problemas encontrados nos espaços urbanos.

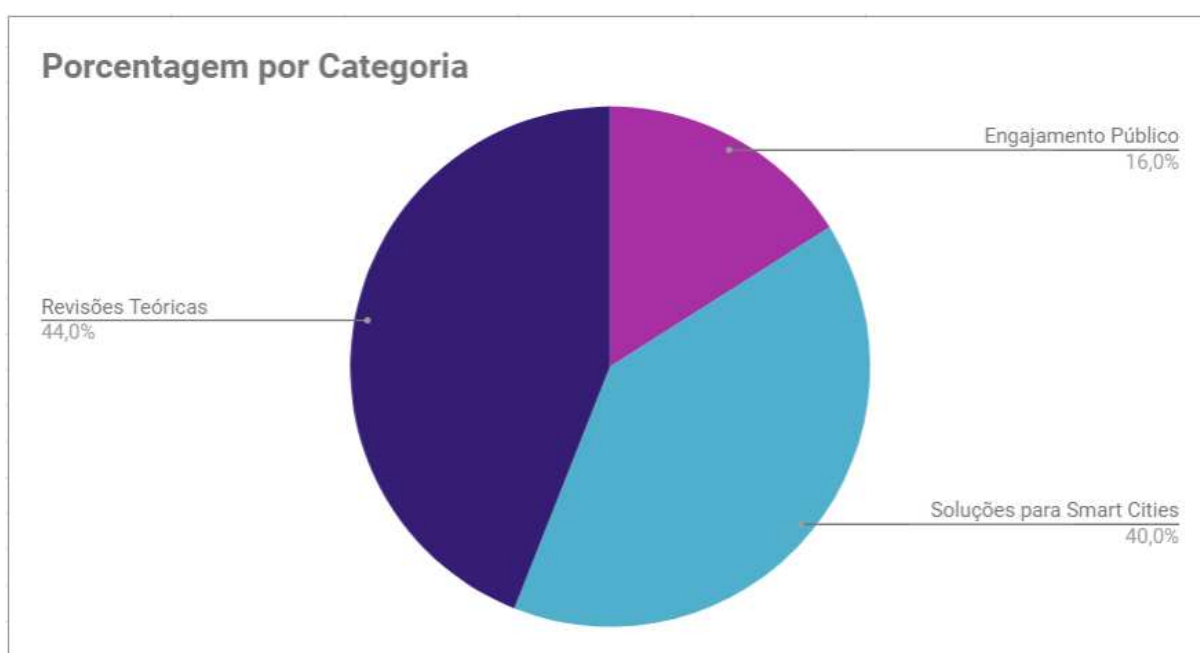
Quatro artigos se aplicam à definição de trabalhos sobre engajamento público. Eles foram agrupados nesta categoria porque são estudos que se referem a soluções para cidades inteligentes que se concentram no aumento e análise do engajamento público e participação em ações de tomada de decisão usando iniciativas de dados abertos.

Na última categoria, soluções para cidades inteligentes, dez trabalhos foram agrupados. O que os caracteriza é o fato de apresentarem propostas mais focadas para a solução de problemas específicos enfrentados por cidades inteligentes. As soluções são aplicadas aos diferentes domínios presentes em cidades inteligentes, tais como: saúde, educação, mobilidade, meio ambiente, cidadania, governança e economia. Todos eles apresentam idéias de como usar dados abertos e discutem as conseqüências de seu uso.

Os domínios das cidades inteligentes utilizados nesta categoria baseiam-se num estudo da Universidade de Viena. Giffinger e Pichler-Milanović (2007), desenvolveram uma métrica de avaliação para classificar cidades européias de tamanho médio. Essa métrica usa indicadores específicos para cada um dos seis principais componentes, também chamados de dimensões, da cidade inteligente: economia inteligente, pessoas inteligentes, governança inteligente, mobilidade inteligente, ambiente inteligente e vida inteligente.

A Figura 3 apresenta a porcentagem por categoria dos artigos selecionados. A porcentagem de revisões teóricas (44,0%) e soluções para cidades inteligentes (40,0%) é quase igual. Trabalhos baseados em engajamento público são apenas 16,0%. Aparentemente, os estudos sobre a participação pública na tomada de decisões e o empoderamento dos cidadãos por meio de dados abertos inteligentes em cidades inteligentes são minoritários. Artigos sobre conceituações, soluções, ferramentas e tecnologias de dados abertos inteligentes estão mais presentes nos repositórios de pesquisa.

Figura 3 – Porcentagem dos artigos por categoria.



Fonte: Autoria própria

2.2 Resultados da Revisão Sistemática

Esta seção apresenta alguns trabalhos correlatos com o contexto da pesquisa, selecionados de acordo com os tópicos que delimitam as revisões e estudos sobre o tema.

Serão apresentados os seguintes temas: cidades inteligentes, dados inteligentes, *linked data*, engajamento público, trabalhos relacionados a modelos ontológicos e tecnologias semânticas capazes de homogeneizar semanticamente dados heterogêneos. Todos os trabalhos apresentados aqui são resultados da revisão sistemática realizada por Belizario e Berardi (2019).

2.2.1 Cidades Inteligentes: Problemas e Oportunidades

Impulsionados pelos problemas e dificuldades que as cidades inteligentes apresentam, Sanseverino, Sanseverino e Anello (2018) pesquisaram conceitos e definições relacionadas a esse tema e comparar suas implementações em diferentes cidades. Eles fizeram uma abordagem de leitura cruzada para o conceito e implementação de cidades inteligentes da China e da Europa. Eles concluíram que a abordagem europeia para o planejamento é consolidada e baseada em uma visão de longo prazo, enquanto a visão chinesa está alcançando a velocidade da urbanização, implantando infraestruturas críticas na maioria dos casos sem uma visão de longo prazo. Eles também apontam que, para o desenvolvimento do conceito de governança inteligente, é importante ter uma estrutura bem definida sobre engajamento civil na tomada de decisões e dados abertos.

Thompson (2016) revisou o conceito de cidades inteligentes em todo o mundo para entender como as diferentes abordagens usadas por pesquisadores e governos influenciam o desenvolvimento da cidade. Seu estudo refere-se à compreensão da conceituação de cidades inteligentes, com o objetivo de entender o que realmente é uma cidade que se enquadra nessa definição.

Mehta e Buch (2015) revisaram os desafios e oportunidades no uso de *Big Data* no contexto de tecnologias semânticas para cidades inteligentes. Eles introduzem e analisam tecnologias que podem suportar o gerenciamento de dados nesse cenário, como: *Resource Description Framework* (RDF), SPARQL e *Ontology Web Language* (OWL).

Movido por análises de dados urbanos por técnicas como a *Association Rule Mining* (ARM), Balasubramani et al. (2016) exploraram ontologias de domínio para filtrar e remover regras que são variações de um conceito mais geral na ontologia, e substituir grupos de regras por uma regra geral com a intenção de reduzir o número de regras iniciais preservando a semântica. Assim, permitindo efetivamente que os administradores usem dados abertos para gerar padrões, use-os para tomar decisões e facilitar a análise dos dados gerados pelo espaço urbano.

O estudo realizado por Moustaka, Vakali e Anthopoulos (2018) foca no uso da ciência de dados em cidades inteligentes. Eles conduziram uma revisão sistemática para identificar os principais tópicos, serviços e métodos usados para monitorar e analisar os dados produzidos pelos centros urbanos. Segundo os autores, os dados abertos são um dos principais pilares quando se trata de governança inteligente. As soluções de dados abertos permitem a expansão da comunicação entre agentes públicos e cidadãos e permitem a melhoria dos serviços.

McMillan et al. (2016) consideram como os dados são produzidos e utilizados nas cidades. Para seu trabalho, foram feitas entrevistas em quatro cidades para entender o papel que os dados desempenham no governo da cidade. Além disso, eles discutem o que esses dados significam para as cidades e seus cidadãos com base em quatro pontos de observação: burocracia, cidadãos, poder e abertura.

Com base no uso de serviços móveis usados pela maioria dos cidadãos de grandes áreas urbanas, Johansson, Lassinantti e Wiberg (2015) estudam os paradigmas da computação móvel com o objetivo de ampliar o ciclo de vida dos serviços móveis por meio do uso de dados abertos e participação ativa do público. O ciclo de vida do consumo de serviços é unidirecional na maioria dos casos, mas os autores viram nos próprios consumidores um potencial para ampliar o uso desses serviços.

Degbelo (2017), realizou um estudo sobre a importância de dados conectados e a visualização de dados em cidades inteligentes, especificamente no aspecto de transparência da informação. Segundo ele, *linked data* é capaz de ajudar na transparência da informação, bem como os impactos da visualização na transparência para os cidadãos. Ainda neste tópico, Piovano et al. (2014) tiveram como objetivo fornecer uma visão geral dos problemas de visualização de dados, analisando os esforços na visualização urbana, especialmente ao considerar duas questões críticas: planejamento urbano e eficiência energética. Eles apresentaram as abordagens mais comuns para as disciplinas de visualização de dados e as contextualizaram nas iniciativas e projetos europeus mais avançados.

Alavi et al. (2018) tentaram prever tendências em cidades inteligentes, eles realizaram uma revisão da Internet das coisas (IoT) em cidades inteligentes. Neste artigo, os autores revisam os paradigmas envolvendo IoT e tecnologias emergentes que podem ser usadas com esse conceito, como *Radio-Frequency Identification* (RFID) e computação em nuvem.

Williamson (2015) se concentra no domínio da educação em cidades inteligentes. Seu estudo visa compreender como o uso de tecnologias e práticas de *big data* pode definir e colaborar com o futuro da educação e com novos métodos de ensino. Ele demonstra como a educação está sendo reinventada em relação a cidades inteligentes que dependem de *big data* para seu funcionamento. Além de discutir como a educação pode ser governada pela produção e circulação de dados e algoritmos.

Gebetsroither-Geringer, Stollnberger e Peters-Anders (2018) apresentam problemas típicos de aplicações web em termos de design e implementação de aplicativos e avaliação de usabilidade descrevendo três aplicações de estudo de caso que foram desenvolvidas. As aplicações têm o objetivo comum de fornecer recursos interativos de geovisualização e análise que são adaptados para apoiar os usuários em seus processos de planejamento urbano usando dados espaciais abertos públicos.

Como conclusão das revisões teóricas, percebeu-se que algumas tecnologias são muito difundidas na literatura quando se trata de dados inteligentes, como *big data* e dados vinculados. Mas outras soluções ainda aparecem quando se trata de dados inteligentes, como mineração de dados, IoT e tecnologias semânticas, que são usados em conjunto com dados vinculados. Além disso, McMillan et al. (2016) enfatiza que a participação pública é importante para o ciclo de vida dos dados das cidades e que existem outros pontos relevantes para discussões como a burocracia para a abertura e uso de dados.

2.2.2 Engajamento Público

Dados abertos são definido por Algemili (2016) como uma abordagem de gerenciamento de dados de forma a possibilitar o fluxo estruturado de informações para aqueles que têm necessidade ou interesse em utilizar essas informações. Ele permite que diferentes tipos de usuários acessem, organizem e usem dados de maneiras que façam sentido para eles. Em seu trabalho, são discutidos os vários desafios que os projetos de dados abertos podem encontrar durante a transformação de conjuntos de dados passivos e privados para a cultura de dados abertos.

Para identificar as questões que cercam as iniciativas de *Open Government Data* e os obstáculos em suas implementações, o autor realizou entrevistas com várias pessoas de comunidades práticas, como equipes de Tecnologia da Informação, gerentes de projeto e executivos de nível médio. Como conclusão do trabalho, Algemili indica doze pontos que devem ser cobertos por políticas de dados abertos, são: integridade dos dados, originalidade, atualização de dados, acessibilidade, legibilidade por máquina, disponibilidade, não-exclusividade, licença livre para uso, gratuidade, segurança e propriedade de dados. Além disso, refere-se aos dados conectados como dados publicados na Web de modo que sejam legíveis por máquina, seu significado é explicitamente definido, está vinculado a outros conjuntos de dados externos e pode, por sua vez, ser vinculado a conjuntos de dados externos. Segundo o autor, os dados vinculados abrangem a rede de dados e interliga fontes descentralizadas e heterogêneas. A web de dados visa substituir dados isolados por um conjunto de dados distribuídos e construídos no topo da arquitetura web.

Balestrini et al. (2017) desenvolveram uma estrutura que pode ser usada por grupos comunitários, organizações e partes interessadas nos governos para orientar e desenvolver processos participativos. Seu objetivo era fornecer uma estrutura que ajudasse as comunidades, os pesquisadores e os conselhos municipais a planejar e executar intervenções sensoriais inovadoras para lidar com questões locais. Eles discutem como o projeto foi usado pelas comunidades em Bristol para medir e monitorar o problema da habitação úmida. E dados abertos foram coletados para ajudar a visualizar a prevalência de casas úmidas e sua correlação com outros fatores.

O trabalho de Hartog e Mulder (2017) baseiam-se nos problemas de comunicação entre o governo e os cidadãos e, portanto, traz uma abordagem diferente sobre como lidar com os problemas das cidades inteligentes. Tentando estabelecer mais confiança e evitar falhas de comunicação nas colaborações entre as administrações públicas e a sociedade participativa, um vocabulário de cidadão foi desenvolvido em um formato *Linked Open Data*. Este vocabulário define relacionamentos, qualidades, mapeamento da estrutura de um domínio de conhecimento e conceitos que são legíveis por máquina usando tecnologias como *Extensible Markup Language (XML)*, *Resource Description Framework (RDF)* e *SPARQL*.

Em uma abordagem semelhante, Dambruch (2018) discute como fornecer feedback

computado sobre as propostas dos cidadãos com base em dados abertos e conhecimento especializado em planejamento urbano e participação pública usando *Domain Specific Language* (DSL). A conclusão é que as DSLs podem ser aplicadas com sucesso para permitir uma nova maneira de acessar dados de uma forma mais conveniente e compreensível, abstraindo-se dos detalhes técnicos e focalizando os aspectos do domínio.

Como conclusão parcial, os autores apresentados acima afirmam que, para que as soluções de dados abertos funcionem, deve haver uma cultura sobre identificação e comunicação de problemas entre governo e cidadãos. Hartog e Mulder (2017) apresentam propostas de dados vinculados com o objetivo de melhorar essa comunicação e Dambruch (2018) demonstra como o conhecimento popular pode contribuir para a geração de dados.

2.2.3 Soluções Aplicadas a Cidades Inteligentes

Segundo Honarvar e Sami (2016), os dados de mineração nas cidades envolvem um grande número de desafios, pois várias fontes de dados em cidades inteligentes geralmente são grandes, independentes, heterogêneas e nenhuma semântica é integrada e anotada a elas. Movidos por esses problemas, eles desenvolveram uma abordagem que aproveita dados abertos vinculados, tecnologias semânticas da Web, mecanismos de mineração de dados e plataformas de processamento de *Big Data*.

Consoli et al. (2015a) apresentam uma plataforma web de relatórios de falhas urbanas para ajudar os cidadãos a relatar problemas urbanos comuns, como falhas de ruas, buracos ou luzes quebradas da rua, e para apoiar a administração pública local na resposta e correção desses problemas. A ferramenta desenvolvida é baseada em um modelo de dados semânticos projetado para a cidade, que integra fontes de dados distintas, oportunamente reprojatadas para atender aos princípios da Web Semântica e dados abertos vinculados.

Com base nos enormes volumes de dados urbanos, Beseiso, Al-Alwani e Altameem (2017) propõem uma estrutura que utiliza os dados brutos da cidade inteligente de vários recursos e os armazena em um banco de dados NoSQL. A estrutura transforma os dados dispersos em dados processáveis por máquina que podem ser úteis para o bem-estar dos cidadãos.

Gastaud (2017) trabalha no desenvolvimento de novos serviços baseados na Internet das Coisas (IoT). O objetivo principal é ter uma rede de sensores que possam interagir facilmente com *softwares* para resolver problemas de cidadãos e cidades.

Segundo Liu et al. (2017), as cidades inteligentes podem aumentar a complexidade do gerenciamento dos dados, devido a seus tamanhos, diversidade e problemas de privacidade. Eles fornecem uma nuvem científica chamada SciCloud, essa plataforma fornece recursos de recursos de computação sob demanda, um gerenciamento de dados escalonável e um ambiente de análise de dados no local para dar suporte à pesquisa científica usando dados de cidades inteligentes.

Iwaniak et al. (2016) afirmam que os documentos de planejamento espacial

contêm informações sobre os princípios e direitos do uso do solo em diferentes zonas de uma autoridade local, mas não são legíveis por máquina. Assim, eles desenvolveram uma ontologia de planejamento espacial para anotar diferentes seções de documentos de planejamento espacial com metadados semânticos no *resource description framework in attributes* (RDFa) para identificar relações entre regulamentos no texto, objetos espaciais nos desenhos e links para recursos externos.

Segundo Consoli et al. (2015b), um bom modelo de dados para cidades inteligentes deve ser capaz de descrever e integrar dados de múltiplos domínios. Movidos por isso, eles desenvolveram uma plataforma semântica para o município de Catania, na Itália. A plataforma é baseada no paradigma *Linked Open Data* e para atingir seu objetivo, uma ontologia foi descrita. Não existe um padrão de referência bem estabelecido para a tecnologia de plataforma IoT, de acordo com Fortino et al. (2018). Para isso, eles desenvolveram alguns padrões de IoT, o projeto INTER-IoT está dando suporte e facilitando para as empresas projetar dispositivos IoT, objetos inteligentes e serviços e levá-los ao mercado rapidamente, criando assim novos ecossistemas interoperáveis de IoT.

Como conclusão dos trabalhos apresentados acima, entendeu-se que os dados semânticos e soluções web semânticas estavam mais presentes do que o esperado na literatura. E que soluções web semânticas em conjunto com *linked open data* são usadas para integrar dados de diferentes fontes e em diferentes formatos em um único vocabulário, assim como este projeto propõe.

2.2.4 Discussão sobre os resultados da Revisão Sistemática

Como conclusão da revisão sistemática pode-se apontar alguns dos principais problemas enfrentados pelas cidades inteligentes que são destacados na literatura. Em um aspecto mais técnico, foi identificado que os principais problemas se referem à quantidade de dados gerados pelos centros urbanos, à qualidade desses dados, à acessibilidade dos mesmos pelos cidadãos, bem como aos governadores e pesquisadores. No entanto, há um aspecto social e cultural deste problema, ainda há dúvidas sobre como exatamente usar esses dados para melhorar a qualidade de vida das pessoas e a qualidade do serviço oferecido por ela. Além disso, alguns autores dizem que o engajamento social pode ser visto como um problema a ser enfrentado, especialmente quando o assunto são dados abertos. Incentivar o uso de dados coletados e fornecidos é um desafio que deve ser abordado por funcionários do governo se eles desejarem que a comunidade pública seja ativa na implementação, melhoria e monitoramento do ambiente urbano. Além da lacuna de comunicação entre o governo e os cidadãos, que não possuem o mesmo vocabulário.

Entre as soluções aplicadas para estruturação e representação destaca-se o uso de soluções *Big Data* para análise e interpretação de dados gerados por cidades. Bem como as soluções *Linked Data* combinadas aos princípios semânticos, com a capacidade de estruturar e conectar dados de diferentes fontes e em diferentes formatos, destacando o uso

de ontologias utilizadas por alguns autores para definir vocabulários, modelos e conceitos dentro de um domínio específico que podem servir as cidades. Assim, pode-se falar sobre a aplicação dos princípios de IoT, usados por alguns autores para conectar dispositivos responsáveis pela coleta, análise e armazenamento de dados em cidades inteligentes.

Descobriu-se que termo *smart data* não é um termo muito difundido e parece não ter um significado estável. Mas pode-se dizer que esse termo se refere aos novos métodos e abordagens empregados pela informação e pelas TICs para lidar, analisar e interpretar grandes quantidades de dados de diferentes domínios e fontes. Na literatura, mais de um conceito é apresentado, mas de forma geral todos se referem a métodos para lidar com dados, seus tratamentos, interpretações, análises e manipulações.

A aplicação de dados abertos pelas cidades pode ajudar na confiança dos serviços prestados, aumentando a transparência e a integridade do setor público. O uso deste conceito também pode permitir uma maior participação e inclusão da comunidade na tomada de decisões, e permitir que os próprios cidadãos possam produzir e usar dados para fins públicos e privados. Além de permitir que eles produzam suas próprias iniciativas com base nos dados abertos fornecidos para resolver ou aliviar problemas presentes nos diferentes domínios das cidades inteligentes.

Esta revisão permitiu concluir que o uso de dados inteligentes e abertos está, de fato, presente no contexto de cidades inteligentes, abordando principalmente questões relativas à integração e análise para gerar informações úteis. Dentre as abordagens presentes na literatura, a mais destacada foi o uso de tecnologias semânticas para tratar e conectar dados de fontes heterogêneas.

3 TRABALHOS CORRELATOS

Após avaliar a necessidade de tecnologias semânticas em cidades inteligentes, percebe-se a importância de apresentar projetos de pesquisa com objetivos e problemáticas semelhantes aos apresentados por este trabalho de conclusão de curso. Assim, nesta seção são apresentados trabalhos correlatos, que abordam a integração de dados em cidades inteligentes, e os domínios de educação e mobilidade em cidades inteligentes.

Nesi et al. (2017) propuseram uma plataforma completa de código aberto para impulsionar a integração, enriquecimento semântico, publicação e exploração de dados públicos para promover cidades inteligentes em administrações locais e nacionais. De acordo com os autores, as tecnologias semânticas da Web e os princípios *Linked Open Data* fornecem um meio de compartilhar conhecimento sobre cidades como sistemas físicos, econômicos, sociais e técnicos, permitindo o desenvolvimento de serviços de cidades inteligentes.

A primeira fase do projeto introduzido consiste na ampla pesquisa das fontes de dados relevantes (fontes locais, regionais, nacionais e internacionais). As fontes de dados públicas e privadas foram analisadas e selecionadas para a plataforma. Ela permite um processo iterativo, para que novas fontes de dados possam ser inseridas ao longo do tempo.

A segunda fase descrita por Nesi et al. (2017) consiste na integração dos dados, por meio de vocabulários descritos por ontologias e ferramentas para o tratamento de dados específicos. Após a integração dos dados, os autores tornaram as informações de valor agregado resultantes do processo de integração públicas e pesquisáveis na Web como *Linked Open Data*.

De forma ampla, o trabalho de Nesi et al. (2017) tem objetivos e uma estrutura técnica muito semelhante à proposta por este trabalho. Porém, existe uma diferença quanto ao vocabulário aplicado e os resultados provenientes da integração. O trabalho descrito utiliza vocabulários para mapear cenários de coleta e armazenagem dos dados, se preocupando mais com seus aspectos físicos e sobre como a disponibilidade dos dados sobre diversas áreas da cidade pode trazer benefícios para os cidadãos.

Este trabalho de conclusão de curso se preocupa com os cenários sociais complexos que existem em cidades inteligentes e sobre como a integração semântica dos domínios pode trazer novas informações relevantes para gestores, cidadãos e pesquisadores. Assim, o foco não se encontra em como os dados são capturados ou armazenados, mas sim sobre como eles estão relacionados e sobre como isto pode mostrar novas formas de lidar com os problemas existentes nas cidades.

O trabalho de Janssen, Matheus e Zuiderwijk (2015) mostra dois estudos de casos, dentre eles um sobre mobilidade envolvendo o uso de *Big Open Linked Data* (BOLD). De acordo com os autores, a cidade do Rio de Janeiro no Brasil desenvolveu uma das maiores

infraestruturas BOLD do mundo. A cidade possui mais de trinta mil conjuntos de dados e sete APIs para dados em tempo real.

A iniciativa do Rio de Janeiro foi criada para solucionar problemas relacionados a transporte público e tráfego. A Prefeitura do Rio entendeu que não apenas a cidade precisa ser inteligente, mas também que os cidadãos devem ser auxiliados a tomar as melhores decisões e se tornar mais inteligentes, a respeito do uso de dados e tecnologias. Assim, foi estabelecida uma parceria com o empresas privadas e aplicativos, que permitem aos cidadãos enviar informações em tempo real sobre a cidade, como condições de trânsito e acidentes (JANSSEN; MATHEUS; ZUIDERWIJK, 2015).

Essa iniciativa pode ainda ser complementada por dados abertos segundo os autores. No entanto, a coleta de dados é difícil devido a alguns motivos: Um grande número de sensores precisa estar disponível para coletar dados; Os sensores são usados para rastrear o tráfego em tempo real; Software e recursos humanos têm que estar especializados em processamento de dados e análises; E Finalmente, é necessário apoio político para obter acesso a dados abertos;

Portanto, o diferencial deste trabalho de conclusão de curso, quando comparado com o trabalho de Janssen, Matheus e Zuiderwijk (2015) se encontra nos dados em que a integração se apoia. A iniciativa da cidade do Rio de Janeiro trouxe benefícios e análises complexas sobre o domínio de mobilidade para a cidade. Entretanto, tem como base dados de sensores e empresas privadas, em maioria, e não realiza a integração semântica entre diferentes domínios da cidades.

De acordo com Romero e Ventura (2007), em ambientes de ensino convencionais, os educadores são capazes de obter *feedback* sobre as experiências de aprendizado dos alunos em interações presenciais com os alunos, permitindo uma avaliação contínua de seus programas de ensino. A tomada de decisão dos processos em sala de aula envolvem observar o comportamento de um aluno, analisar dados históricos e estimar a eficácia das estratégias pedagógicas. No entanto, quando os alunos trabalham em ambientes eletrônicos, esse monitoramento informal não é possível, os educadores devem procurar outras maneiras de obter essas informações.

Movidos por esses problemas, os autores propõem uma pesquisa sobre *data mining* na área de educação entre os anos de 1995 a 2005, avaliando como a a mineração de dados e informações podem ajudar os educadores e sobre quais técnicas podem ser utilizadas para esse propósito, principalmente no que diz respeito à educação a distância.

Por este motivo esta pesquisa está sendo apontada como um trabalho correlato, ela investiga quais dados são relevantes para a educação e quais são as dificuldades e obstáculos em sua obtenção, além de mostrar possíveis técnicas de mineração de dados dessa área. Dentre as técnicas, os autores apontam: mineração de texto, mineração *Web*, *clustering* e mineração sequencial de padrões.

O trabalho de Auceli et al. (2019) teve o objetivo de analisar se existe uma relação

entre o rendimento de alunos com a quantidade de ocorrências policiais do bairro em que se encontram a partir da integração semântica de bases de dados abertas heterogêneas dos domínios de educação e segurança. Os dados utilizados são referentes à cidade de Curitiba, no estado do Paraná, Brasil. Para realizar a integração uma ontologia foi proposta com o intuito de unificar semanticamente o domínio das bases e especificar formalmente o relacionamento dos dados. Segundo os autores mencionados, o experimento realizado teve resultados suficientes, uma vez que foi possível obter uma análise com base nos dados integrados entre os domínios de segurança pública e educação. Porém vale ressaltar que a análise foi feita com uma porção dos dados disponíveis pelas bases de dados. E que foram observadas dificuldades com relação às bases de dados, uma vez que é necessária uma limpeza dos dados antes da utilização desses.

4 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados, contextualizados e discutidos os principais conceitos que apoiam o desenvolvimento deste trabalho. Antes de expor a metodologia e desenvolvimento, é necessário desenvolver um arcabouço de conhecimentos e informações sobre as questões que serão apresentadas a seguir. São eles: Cidades inteligentes, mobilidade urbana e educação, *linked data*, *open data* e modelos ontológicos.

4.1 Cidades Inteligentes

As cidades do século XXI são sistemas complexos que são caracterizados por um grande número de cidadãos interconectados, empresas, diferentes meios de transporte, redes de comunicação, serviços e utilidades. O crescimento populacional e o aumento da urbanização levantam uma variedade de problemas técnicos, sociais, econômicos e organizacionais que tendem a comprometer a sustentabilidade econômica e ambiental das cidades (NEIROTTI et al., 2014).

Neste contexto, surge um debate sobre a forma como as novas soluções baseadas em tecnologia, bem como novas abordagens para o planejamento urbano e a vida, podem assegurar a viabilidade e prosperidade em áreas metropolitanas (ALAWADHI et al., 2012).

Nesta discussão, o conceito de cidades inteligentes, utilizado neste trabalho também como cidades inteligentes, tem sido objeto de crescente atenção e aparece agora como um novo paradigma de desenvolvimento urbano inteligente e crescimento socioeconômico sustentável, cuja origem pode ser atribuída a movimentos do final dos anos 90 (HARRISON; DONNELLY, 2011).

Neirotti et al. (2014) dizem no entanto que, apesar do aumento do debate sobre cidades inteligentes pelos planejadores urbanos, a difusão de iniciativas de cidades inteligentes em países com diferentes necessidades e condições contextuais dificulta a identificação de definições compartilhadas e tendências atuais comuns em escala global. Ainda não existe, de fato, um consenso geral sobre o significado do termo cidade inteligente ou sobre quais são os seus atributos descritivos. No entanto, o autor diz que há um consenso sobre o fato de que as cidades inteligentes são caracterizadas por um uso difundido das tecnologias de informação e comunicação, que, em vários domínios urbanos, ajudam as cidades a usar melhor seus recursos.

No entanto, soluções baseadas em TICs podem ser consideradas apenas como um dos vários recursos de entrada para projetos e abordagens de planejamento urbano e de vida que têm como objetivo melhorar a sustentabilidade econômica, social e ambiental de uma cidade. Isto implica que as cidades que estão mais equipadas com sistemas de TICs não são necessariamente cidades melhores, e que o número de iniciativas “inteligentes”

lançadas por um município não é um indicador do desempenho da cidade, mas poderia resultar em uma saída intermediária que reflete a esforços envidados para melhorar a qualidade de vida dos cidadãos (NEIROTTI et al., 2014).

Na Tabela 2 são apresentadas algumas definições para cidades inteligentes encontradas na literatura. Por mais que o foco das formalizações seja diferente, de forma geral todos se referem a TICs como auxiliadoras e suportes em diferentes áreas das cidades para melhorar a qualidade de vida dos seus cidadãos.

Tabela 2 – Definições de Cidades Inteligentes

Referência	Definição
Hall et al. (2000)	"Uma cidade que monitora e integra as condições de todas as suas infra-estruturas críticas, incluindo estradas, pontes, túneis, ferrovias / metrô, aeroportos, portos, meios de comunicações, água, energia, até mesmo grandes edifícios, pode otimizar seus recursos, planejar suas atividades de manutenção preventiva e monitorar os aspectos de segurança, maximizando os serviços para seus cidadãos."
Giffinger e Pichler-Milanović (2007)	"Uma cidade que tem um bom desempenho na economia, nas pessoas, na governança, na mobilidade, no meio ambiente e na vida, construída com base na combinação inteligente de recursos e atividades de cidadãos autoconfiantes, independentes e conscientes."
Washburn et al. (2009)	"O uso de tecnologias <i>Smart Computing</i> para tornar os componentes e serviços de infraestrutura críticas de uma cidade, que incluem administração municipal, educação, saúde, segurança pública, imóveis, transporte e serviços públicos, mais inteligentes, interconectados e eficientes."
Rios (2012)	"Uma Cidade Inteligente é uma cidade que dá inspiração, compartilha cultura, conhecimento e vida, uma cidade que motiva seus habitantes a criar e fluir em suas próprias vidas. Uma cidade inteligente é uma cidade admirada, uma embarcação para a inteligência, mas, em última instância, uma incubadora de espaços capacitados."

Neste trabalho será adotado o conceito apresentado por Giffinger e Pichler-Milanović (2007), que concerne a diferentes domínios para a cidade e que abrange tanto o campo técnico quanto sócio-cultural dos desdobramento do uso da tecnologia na vida ativa das cidades e seus cidadãos. Giffinger e Pichler-Milanović (2007) apresentam os domínios das cidades inteligentes, que podem ser interpretadas como grandes áreas pelas quais as cidades são avaliadas e desenvolvidas, e todos os fatores inseridos e tratados por cada um deles:

1. Economia Inteligente (Competitividade): Engloba espírito inovador, empreendedorismo, imagem econômica e marcas, produtividade, flexibilidade do mercado de trabalho, inserção internacional e capacidade de transformação.
2. Pessoas Inteligentes (Capital Social e Humano): Se refere a nível de qualificação, afinidade à aprendizagem ao longo da vida, pluralidade social e étnica, flexibilidade, criatividade, cosmopolitismo, participação na vida pública.
3. Governança Inteligente (Participação): Envolve a participação na tomada de decisões, serviços públicos e sociais, governança transparente e estratégias e perspectivas políticas.
4. Mobilidade Inteligente (Transportes e TIC): Trata da acessibilidade local, (inter) acessibilidade nacional, disponibilidade de infraestruturas TICs, sistemas de transportes sustentáveis, inovadores e seguros.
5. Ambiente Inteligente (Recursos naturais): Se preocupa com uma atratividade das condições naturais, poluição, proteção ambiental, manejo sustentável de recursos.
6. Vida Inteligente (Qualidade de vida): Envolve recursos culturais, condições de saúde, segurança individual, qualidade da moradia, instalações educacionais, atratividade turística e coesão social.

Além dos domínios já apresentados, Nam e Pardo (2011) identificam três áreas de componentes e fatores importantes para a implementação de cidades inteligentes: tecnologia (infra-estrutura de hardware e software), pessoas (criatividade, diversidade e educação) e instituição (governança e política). Segundo os autores, dada a existência da conexão entre os fatores, uma cidade é inteligente quando os investimentos em capital humano/social e infraestrutura de tecnologias da informação incentivam o crescimento sustentável e melhoram a qualidade de vida, por meio de uma governança participativa.

4.2 Mobilidade Urbana e Educação

Como já dito anteriormente neste trabalho, relacionar educação e mobilidade urbana é um trabalho árduo, porém possível. Antes é importante realizar apontamentos sobre exclusão social e os aspectos presentes nessa discussão.

Stanley et al. (2007) apontaram que, enquanto a pobreza é vista como a diferença entre a quantidade de renda necessária para sustentar um indivíduo ou família dentro de seu ambiente de vida, a exclusão social é vista como um conceito mais abrangente. O termo

exclusão social, embora ainda muito dependente de medidas de renda, também reconhece que podem haver outras barreiras que dificultam a participação plena das pessoas na sociedade. Essas barreiras podem incluir falta de emprego, moradia adequada, educação, saúde e transporte.

Antes de relacionar mobilidade urbana e educação é importante compreender o significado dos dois termos isolados. Os índices de mobilidade são difíceis de definir e quantificar, mas pesquisadores incluem dimensões como causa, distância, quantidade, tempo e localização. Além disso, há uma distinção entre mobilidade escolar e mobilidade residencial, pois uma mudança de endereço não requer uma mudança de escola e vice-versa. Já o desempenho escolar de um aluno é uma variável mais constante, já que pode ser medido por testes de aproveitamento, notas e progresso de idade-série, por exemplo (HEINLEIN; SHINN, 2000).

A mobilidade urbana é composta por várias modais, como transporte público, qualidade das vias e ciclovias, por exemplo. O foco deste trabalho de conclusão de curso se encontra na análise sobre a acessibilidade à transporte público.

As crianças que moram em lares caracterizados por múltiplos tipos de desvantagens socioeconômicas foram consideradas mais vulneráveis aos efeitos de eventos estressantes da vida, como a desistência escolar, em comparação com crianças que sofrem menos desvantagens. E dentre essas desvantagens pode-se avaliar a distância e a dificuldades para chegarem às suas escolas (DUBOIS et al., 1994).

Raychaudhuri et al. (2010) constatou que inúmeros estudos foram feitos para identificar os fatores que podem afetar o desempenho acadêmico de alunos. O desempenho acadêmico dos alunos depende de uma série de fatores socioeconômicos, como a frequência dos alunos na classe, renda familiar, educação de mães e pais, proporção professor-aluno, presença de professor treinado na escola, sexo do aluno e distância das escolas.

A mobilidade geográfica é uma influência diretamente relacionada e negativa ao desempenho acadêmico dos alunos. Esses efeitos aversivos são mais notáveis nas populações mais vulneráveis e persistem mesmo sob tentativas de controlar o status socioeconômico. Em particular, parece que a característica presente em grupos mais vulneráveis e propensos à instabilidade acadêmica e desistências é a distância das suas escolas e dificuldade sobre a mobilidade geográfica (INGERSOLL; SCAMMAN; ECKERLING, 1989).

4.3 *Linked Data*

Sob uma perspectiva histórica, a *World Wide Web* alterou a maneira como compartilha-se conhecimento, diminuindo a barreira para publicar e acessar documentos como parte de um espaço global de informações. Os *links* de hipertexto permitem que os usuários percorram esse espaço de informações usando navegadores da Web, enquanto os mecanismos de pesquisa indexam os documentos e analisam a estrutura de *links* entre eles para inferir relevância potencial para as consultas de pesquisa dos usuários (BRIN; PAGE,

1998).

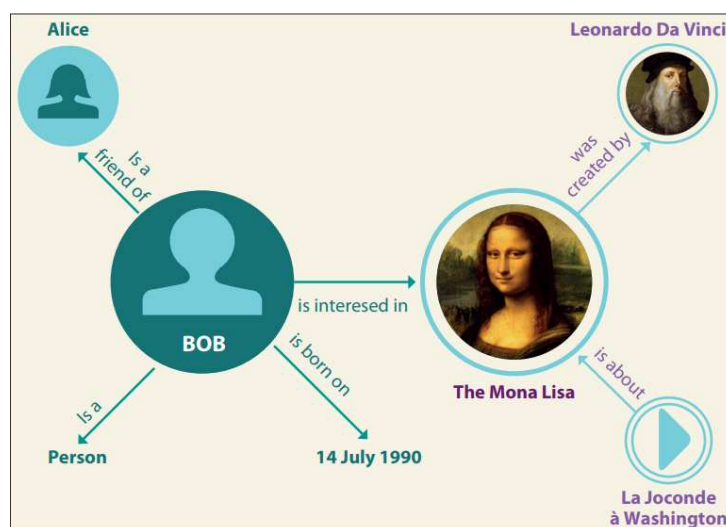
Apesar dos benefícios indiscutíveis que a Web oferece, tradicionalmente os dados publicados na Web são disponibilizados como material bruto em formatos como *Comma-separated values*(CSV) ou *Extensible Markup Language*(XML), ou marcados como tabelas *HyperText Markup Language*(HTML), sacrificando muito de sua estrutura e semântica. Na Web de hipertexto convencional, a natureza da relação entre dois documentos vinculados é implícita, pois o formato de dados, ou seja, HTML, não é suficientemente expressivo para permitir que entidades individuais descritas em um documento particular sejam conectadas por *links* tipados a entidades relacionadas (BIZER; HEATH; BERNERS-LEE, 2011).

Segundo Klyne e Carroll (2006), enquanto as unidades principais da Web de hipertexto são documentos HTML conectados por *hiperlinks*, as iniciativas *Linked Data* confiam em documentos que contêm dados no formato *Resource Description Framework* (RDF). No entanto, em vez de simplesmente conectar esses documentos, as práticas *Linked Data* usam o RDF para fazer declarações explícitas que vinculam conceitos e objetos no mundo.

Resource Description Framework (RDF) é uma série de especificações criadas pela *World Wide Web Consortium* (W3C), com a qual é possível representar um domínio a partir construção de seus conhecimentos em partes e da definição de regras sobre o significado ou a semântica dessas partes.

A Figura 4 apresenta um exemplo de triplas RDF, neste exemplo há seis triplas que representam a relação entre recursos diferentes. Na relação <sujeito> <predicado> <objeto>, o sujeito e o objeto representam dois recursos que são relacionados por um predicado.

Figura 4 – Cenário descrevendo um conjunto de triplas e seus relacionamentos.



No exemplo de Isotani e Bittencourt (2015), $\langle \text{Bob} \rangle \langle \text{is a friend of} \rangle \langle \text{Alice} \rangle$, $\langle \text{Bob} \rangle$ representa um sujeito, que é um recurso, $\langle \text{Alice} \rangle$ representa o objeto que é outro recurso e $\langle \text{is a friend of} \rangle$ representa uma propriedade que relaciona $\langle \text{Bob} \rangle$ à $\langle \text{Alice} \rangle$. Da mesma forma $\langle \text{A Mona Lisa} \rangle \langle \text{was created by} \rangle \langle \text{Leonardo Da Vinci} \rangle$ representa a relação entre o sujeito $\langle \text{A Mona Lisa} \rangle$ e o objeto $\langle \text{Leonardo Da Vinci} \rangle$, por meio da relação/propriedade $\langle \text{was created by} \rangle$.

De acordo com Davies, Studer e Warren (2006), o objetivo das tecnologias semânticas é oferecer novas abordagens para o gerenciamento de informações e processos, cujo princípio fundamental é a criação e uso de metadados semânticos. Seu uso é importante para integrar informações de fontes heterogêneas, por exemplo.

O resultado, ao qual define-se a *Web of Data*, pode ser mais precisamente descrito como uma rede de conceitos no mundo, descrita por dados na Web. Para consumir dados em formato *Linked Data*, é necessário utilizar uma linguagem específica, o SPARQL. O padrão SPARQL é uma linguagem de consulta amplamente usada para consultar dados em padrão RDF e é implementada pelos principais repositórios RDF (SEABORNE; PRUD'HOMMEAUX, 2008).

Formalizando o conceito, Bizer, Heath e Berners-Lee (2011) dizem que os dados conectados se referem sobre o uso da Web para criar links entre dados de diferentes origens. Esses podem ser tão diversos quanto os bancos de dados mantidos por duas organizações em diferentes localizações geográficas, ou simplesmente sistemas heterogêneos dentro de uma organização que, historicamente, não são facilmente interoperáveis no nível de dados. Tecnicamente, dados conectados são dados publicados na Web de forma que sejam legíveis por máquina, que seu significado seja definido explicitamente, que estejam vinculados a outros conjuntos de dados externos e que pode, por sua vez, ser conectados a mais conjuntos de dados externos.

Há um conjunto de práticas recomendadas para publicar e interligar dados estruturados na Web. Essas práticas recomendadas foram introduzidas por Berners-Lee (2006) em sua arquitetura de dados do *Linked Data* e tornaram-se conhecidas como os princípios do *Linked Data*, adotados pela W3C como padrões de publicação de dados conectados. Esses princípios servem para publicar dados na Web de forma que todos os dados publicados se tornem parte de um único espaço de dados global. São eles: Use *Universal Resource Identifier* (URI) como nomes para as coisas; Use HTTP URIs para que as pessoas possam procurar esses nomes; Quando alguém procura um URI, forneça informações úteis, usando os padrões (RDF, SPARQL); Inclua links para outros URIs, para que eles possam descobrir mais coisas;

Conforme os autores, Isotani e Bittencourt (2015), a riqueza desta representação permite que um computador possa interpretar os dados representados por estas triplas. Contudo, para que isso seja completamente realizado, é necessário garantir que cada um dos elementos do grafo seja representado e referenciado de maneira única. E isso pode ser

realizado utilizando os URIs.

4.4 Modelos Ontológicos

Segundo Guarino (1997) o termo ontologia surgiu na filosofia como o estudo do “ser enquanto ser”, com o objetivo de estudar a natureza do ser e a estrutura da realidade. Na computação, a pesquisa em ontologias se estendeu por toda a comunidade, e o conceito de ontologias tem sido aplicado em diferentes áreas da computação.

A Web Semântica depende fortemente de ontologias formais para estruturar dados para uma compreensão abrangente e transportável por máquinas. Assim, a criação de ontologias influencia o sucesso da Web Semântica. Elas servem como esquemas de metadados, fornecendo um vocabulário controlado de conceitos, cada um com uma semântica explicitamente definida e processável por máquina. Ao definir modelos de domínio comuns e compartilhadas, as ontologias ajudam as pessoas e as máquinas a se comunicarem de forma concisa, suportando a troca semântica e não apenas a sintaxe (MAEDCHE; STAAB, 2001). Nesse contexto, as ontologias computacionais são um meio de modelar formalmente a estrutura de um sistema, isto é, as classes e relações relevantes que emergem de sua observação e que são úteis para determinados propósitos (GUARINO; OBERLE; STAAB, 2009).

O conceito de modelos ontológicos adotado por este trabalho segue as observações de Guarino, Oberle e Staab (2009), apresentada anteriormente, e as definições de Gruber (1993), que formaliza ontologia como uma especificação explícita e formal de uma conceitualização compartilhada. E entende-se por conceitualização os conceitos, objetos, entidades e relacionamentos presentes em um determinado domínio.

4.4.1 *Resource Description Framework*

A *Ontology Web Language* (OWL) é uma linguagem que amplifica as possibilidades do RDF e do RDFS. Ela oferece uma série de relações para serem declaradas sobre as triplas e diversos construtores que permitem a construção de triplas mais complexas. Sua principal base vem das *Description Logics* (DLs), linguagens de representação de conhecimento utilizadas na construção de ontologias. Elas são lógicas e possuem uma semântica formal, uma especificação precisa do significado. A semântica formal permite que humanos e máquinas possam compreender as ontologias sem riscos de causar ambiguidade quanto ao seu significado, além de permitir a utilização de dedução lógica para inferir informações adicionais dos fatos explicitados na ontologia (ISOTANI; BITTENCOURT, 2015).

Arquivos RDF são modelos que utilizam vocabulários baseados em URIs e sintaxes baseadas em XML. O modelo RDF é formado pelo conjunto de declarações predicado, sujeito e objeto, essa combinação é conhecida como triplas. O objeto e o sujeito são recursos

sobre os quais a RDF permite fazer afirmações. Uma afirmação na RDF expressa a relação do sujeito com o objeto, sendo que o predicado representa a natureza da relação, e é chamada em RDF de propriedade (KLYNE; CARROLL, 2006).

Ainda segundo Isotani e Bittencourt (2015), *Resource Description Framework Schema* (RDFS) é um vocabulário que complementa a RDF e permite especificar características que agregam semântica aos dados. Com ela é possível especificar que uma URI é uma propriedade de um recurso ou que um recurso pertence a uma determinada classe. Esse modelo utiliza o conceito de classes para montar categorias sobre os recursos, podendo até mesmo montar hierarquias de classes, subclasses, propriedades e sub propriedades. Além de permitir restrições de tipos sobre os sujeitos e objetos das triplas, através da especificação de domínios e contradomínios para cada um dos tipos. Essas propriedades da RDFS permitem que se faça inferências sobre os dados.

4.5 Dados Abertos

A disponibilidade de dados abertos cresce, com pressão sendo colocada em todos os tipos de organizações públicas para liberar seus dados brutos. Algumas motivações principais são de que o acesso aberto a dados financiados por fundos públicos proporciona maiores retornos do investimento público, pode gerar riqueza através do uso de produtos, fornece aos gestores políticos, cidadãos e pesquisadores os dados necessários para resolver problemas urbanos complexos (ARZBERGER et al., 2004).

Dados abertos podem ser usados para o desenvolvimento de políticas públicas e prestação de serviços. Pode-se defini-los como dados não restritos à privacidade e não confidenciais que são produzidos com dinheiro público e são disponibilizados sem quaisquer restrições sobre seu uso ou distribuição. Dados privados, confidenciais e classificados são excluídos, pois esse tipo de dado é inadequado para divulgação. Contudo, os dados podem ser fornecidos por organizações públicas e privadas, pois a essência é que os dados sejam financiados por dinheiro público (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

Em síntese, dados abertos são capazes de corrigir a tradicional separação entre organizações públicas e usuários e a abertura de dados leva a duas suposições importantes sobre o governo. Em primeiro lugar, leva a uma suposição sobre a prontidão dos órgãos públicos para um processo de abertura que considera as influências, os discursos e as trocas como construtivas e acolhe opiniões e insumos opostos. Em segundo lugar, isso leva a uma suposição de que o governo deve abandonar o controle, ao menos em parte, exigindo transformações consideráveis do setor público (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

Berners-Lee (2006) sugeriu um esquema de implementação de *open data* chamado de "5 estrelas para Dados Abertos". A Tabela 3 mostra como categorizar um documento na *Web* de acordo com este padrão.

Tabela 3 – 5 estrelas para Dados Abertos Conectados

Estrelas	Atributos
1	Torne os recursos disponíveis na Web, independente do formato sob uma licença aberta
2	Torne os recursos disponíveis como dados estruturados (ex. excel)
3	Utilize formatos não-proprietários (ex. CSV)
4	Utilize URIs para identificar recursos. Isso vai ajudar as pessoas a apontarem para eles
5	Conecte seus dados com dados de outras pessoas para prover contexto (<i>linked data</i>)

Davies (2010) aborda o contexto governamental e explica que explorar o uso de Dados Abertos Governamentais (OGD) pode proporcionar aos indivíduos maior compreensão do estado, e o acesso direto a fatos e informações pode capacitar os indivíduos em suas interações com o governo. A ideia de que “o cidadão informado é o cidadão empoderado” é intuitivamente plausível, e é remover as desigualdades de informação entre os cidadãos e entre cidadão e estado, a fim de obter uma maior governança democrática (DAHL, 2000).

Segundo Isotani e Bittencourt (2015), essa visão de dados abertos governamentais passou a ser fortalecida em 2008, após um memorando do presidente Barack Obama sobre transparência e dados governamentais e pela criação do Data.gov, com o objetivo de disponibilizar um portal de dados abertos no qual os dados do governo norte-americano poderiam ser acessados na Internet por qualquer cidadão. Nessa perspectiva, o governo brasileiro foi um dos fundadores da *Open Government Partnership*, criada em 2011, que conta atualmente com a participação de 65 países. O governo brasileiro criou também o dados.gov.br, portal que disponibiliza dados governamentais.

A Lei de Acesso à Informação (LAI), que entrou em vigor no Brasil em maio de 2012, devido a iniciativa do governo estadunidense, pode ser vista como arcabouço jurídico para a obrigação governamental de disponibilizar dados abertos no Brasil (ISOTANI; BITTENCOURT, 2015).

Sobre a cidade de Curitiba, Nam e Pardo (2011) explicam que o *Intelligent Community Forum* (ICF) anuncia anualmente as cidades premiadas com o título de *Smart21 Communities*, que são as cidades que foram bem sucedidas em termos de cinco fatores referentes a cidades inteligentes. E dentre as cidades que receberam essa premiação entre os anos de 2007 e 2011 pelo ICF se encontra Curitiba, confirmando seu título de cidade inteligente.

Janssen, Charalabidis e Zuiderwijk (2012) ampliam o pensamento de que em vez

de reforçar os processos atuais, os dados abertos devem resultar em um governo aberto, no qual o governo age como um sistema transparente e interage com seu ambiente. Não apenas os dados devem ser publicados, mas devem ser ativamente buscados para conhecimento sobre como melhorar o governo. A divulgação de dados poderia ter efeitos de longo alcance no setor público. Mecanismos para monitorar e responder às perguntas feitas pelo público são, portanto, necessários, e o governo deve ser visto como um sistema aberto que interage com seu ambiente.

Kozievitch et al. (2016) reforçam que a cidade também participa da iniciativa de dados abertos, por meio de diversos atores governamentais, como o Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) e o Governo Municipal de Curitiba, que oferece uma plataforma de distribuição de dados abertos sobre diferentes domínios da cidade. A questão é que os dados disponibilizados pelo estes agentes não são conectados e não têm uma visão integrada sobre os diferentes domínios da cidade.

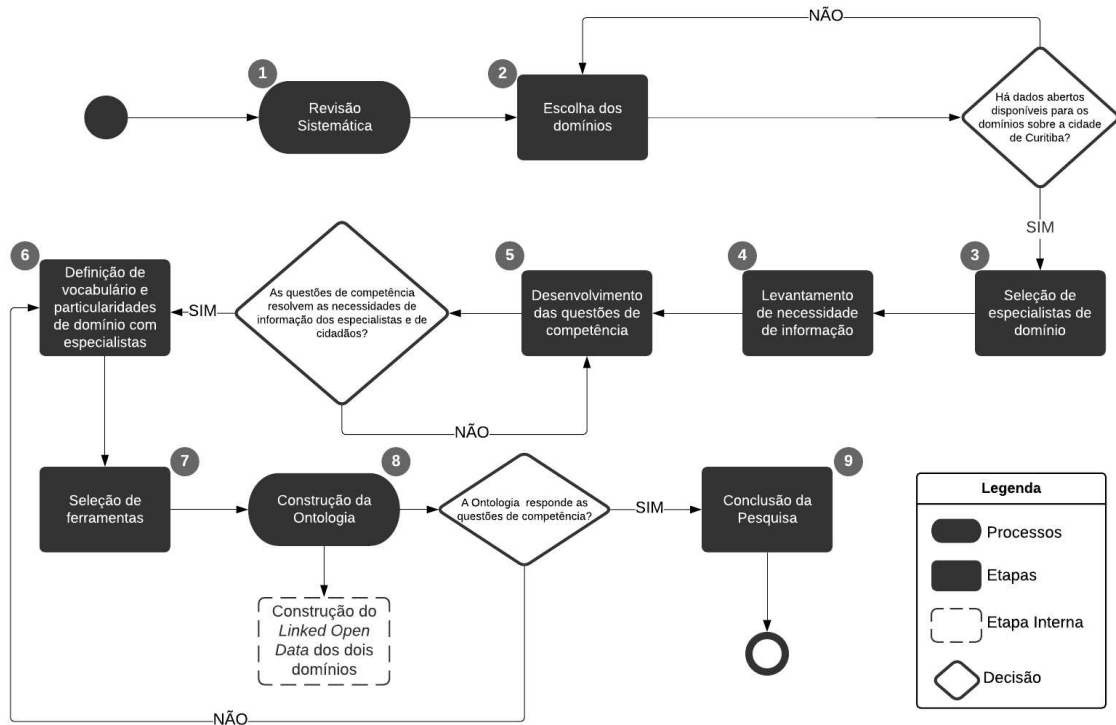
5 METODOLOGIA

Este capítulo apresenta a metodologia proposta por este projeto para alcançar os resultados e os objetivos propostos por este trabalho. A metodologia aplicada será feita em etapas que foram seguidas em ordem durante o processo.

5.1 Etapas de Desenvolvimento

A Figura 5 representa um fluxo das atividades da metodologia de forma resumida. Cada uma das etapas será descrita adiante, seguindo o número exposto na figura.

Figura 5 – Metodologia do estudo



Fonte: Autoria própria

1. Se refere ao processo de revisão sistemática da literatura realizada e descrito no capítulo 2, intitulado Estado da Arte.
2. Engloba a escolha dos domínios a serem estudados. Os domínios foram previamente escolhidos pelos argumentos da subseção 1.1, "Problema e Objeto de Estudo". Os domínios selecionados foram mobilidade urbana e educação, devidamente justificados no subcapítulo 4.2, "Mobilidade Urbana e Educação".

3. Selecionar especialistas de domínio que possam ajudar a definir as necessidades que o modelo ontológico deve ser capaz de cobrir. Serão especialistas de domínio da área de educação e docentes da área de arquitetura e urbanismo.
4. São levantadas as necessidades de informações sobre os dois domínios escolhidos na etapa número 2. Para isso será desenvolvido um instrumento de coleta e entrevistas para compreender como o município lida, interpreta e relaciona os dados coletados desses domínios da cidade.
5. Com as necessidades definidas e especialistas de domínio acompanhando o processo, é possível desenvolver as questões de competência que podem guiar o desenvolvimento e construção da ontologia. O desenvolvimento dessas questões pode ser visto como um recorte da metodologia de Noy, McGuinness et al. (2001). Caso as questões desenvolvidas não sejam capazes de relacionar os dois domínios selecionados anteriormente, elas deverão ser desenvolvidas novamente até serem capazes de guiar um modelo que seja capaz de auxiliar gestores, especialistas e cidadãos.
6. Com as questões de competência bem estabelecidas, o vocabulário da ontologia será definido de acordo com as observações realizadas por especialistas de domínio. Será realizado desta maneira para garantir que a ontologia respeite o escopo, as limitações e particularidades de cada domínio e suas relações.
7. Antes de passar para a construção técnica da ontologia é necessário definir e selecionar as ferramentas necessárias.
8. A fase de construção da ontologia é um processo mais complexo, ela seguirá a metodologia proposta por Noy, McGuinness et al. (2001), definida como "Ontology development 101".
9. Caso a ontologia responda às questões de competência, será possível passar para a última etapa do trabalho. A conclusão do estudo consiste em analisar os desdobramentos do relacionamento semântico entre os domínios selecionados e identificar como um domínio é capaz de impactar em outro.

Por meio deste processo, há a intenção de se alcançar resultados que possam indicar como cidades inteligentes podem conectar seus dados e tomar decisões que levem em consideração as particularidades de seus ambientes, a interação entre a população e os diferentes domínios e espaços que compõem a cidade.

6 DESENVOLVIMENTO

Este capítulo descreve o desenvolvimento das seguintes etapas: Seleção de especialistas de domínio, levantamento de necessidades de informação, desenvolvimento das questões de competência, definição de vocabulário, seleção de ferramentas e construção da ontologia. A fase de conclusão da metodologia será discutida no capítulo 7. Os resultados da fase de desenvolvimento deste trabalho de conclusão de curso estão presentes no *GitHub* e podem ser acessados através do *link*: https://github.com/MateusBelizario/ontology_mobed_2020.

6.1 Seleção de Especialistas de Domínio

Devido à indisponibilidade de agenda dos contatos do IPPUC, optou-se por contar com a colaboração de um especialista de domínio de mobilidade da própria UTFPR. A professora Tatiana Gadda, docente do departamento acadêmico de arquitetura e urbanismo da Universidade Tecnológica Federal do Paraná, com quem foram realizadas entrevistas e discussões presenciais para entender as relações entre mobilidade urbana e educação em Curitiba.

Para o domínio da educação, foi desenvolvido um formulário *online* através da ferramenta *Google Forms*¹. Esse formulário foi enviado para docentes do ensino básico, médio e superior que já haviam participado anteriormente de projetos de extensão da universidade, realizados pelo grupo Emíli@s – Armação em Bits². O objetivo desse formulário é compreender a perspectiva de educadores sobre a importância da mobilidade no desempenho acadêmico dos discentes.

6.2 Levantamento de Necessidades de Informação

O Apêndice A deste trabalho apresenta o instrumento de coleta desenvolvido, o formulário *online*. As necessidades de informação contemplam os especialistas de domínio, porém uma necessidade de informação já percebida foi "como os estudantes" usam o transporte, então o instrumento de coleta contemplou também estudantes da cidade.

A seção direcionada para o perfil de estudante foi desenvolvida para compreender como estudantes de diferentes níveis fazem para se locomover até as instituições de ensino que frequentam, suas percepções sobre o transporte público da cidade de Curitiba e quais vulnerabilidades relacionadas a transporte público já afetaram seu desempenho acadêmico.

A seção desenvolvida para o perfil de especialistas do domínio de educação tem o objetivo de verificar a percepção de educadores e pesquisadores da área sobre a influência da mobilidade urbana na educação e no desempenho acadêmico dos estudantes. Assim

¹Disponível em <https://www.google.com/forms/about/>

²Mais informações em <http://emilias.dainf.ct.utfpr.edu.br/>

como identificar as vulnerabilidades do sistema educacional causadas pela estrutura e situação do transporte público.

A terceira seção se refere aos especialistas de mobilidade urbana, as questões foram desenvolvidas para entender as particularidades da mobilidade, do transporte público da cidade e a relação entre a utilização das linhas e terminais pelos estudantes de Curitiba. Além dos formulários, também foi realizada entrevista presencial com a professora Tatiana Gadda.

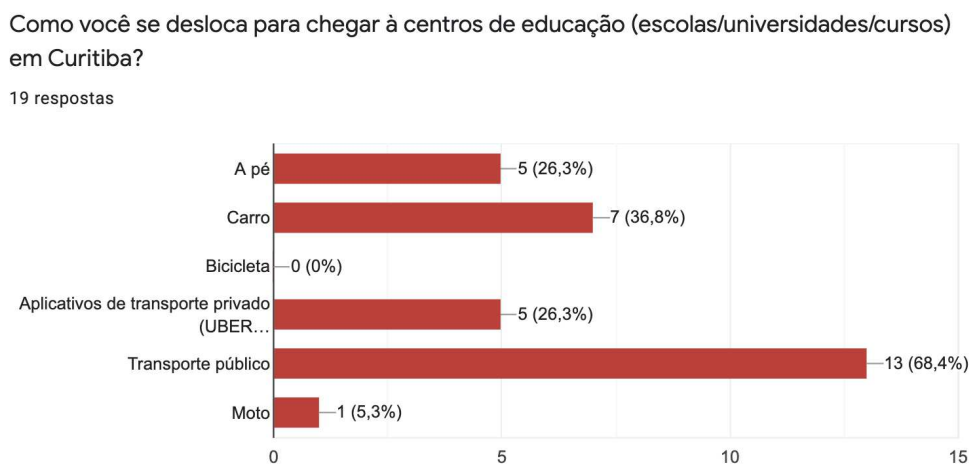
6.2.1 Estudantes

A divisão de perguntas do formulário dirigidas aos estudantes contém quatorze questões. Esta subseção apresenta as respostas para as principais perguntas, as conclusões e observações realizadas. O formulário foi disponibilizado pelas redes sociais, em grupos específicos de estudantes da UTFPR por um mês.

Dezenove estudantes responderam o formulário, sendo todos alunos do ensino superior. A Figura 6 mostra um gráfico de como estes estudantes se deslocam para chegar em suas instituições de ensino. Treze (68,4%) dos estudantes, que responderam o formulário, utilizam o transporte público para chegar a suas universidades, sete (36,8%) utilizam carros e o restante utilizam outros veículos.

A Figura 7 mostra a o nível de importância dada por estes alunos ao transporte público na sua formação acadêmica, sendo o nível cinco muito importante e o nível um nada importante. Dezoito (94,7%) estudantes concederam o nível cinco e apenas um (5,3%) considerou o nível quatro.

Figura 6 – Transporte Utilizados por Estudantes

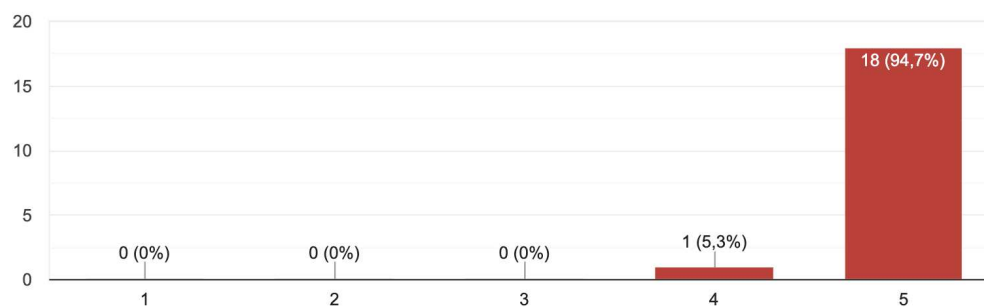


Fonte: Autoria própria

Figura 7 – Importância do Transporte Público

Quão importante considera o transporte público para a Mobilidade de estudantes do ensino Superior?

19 respostas

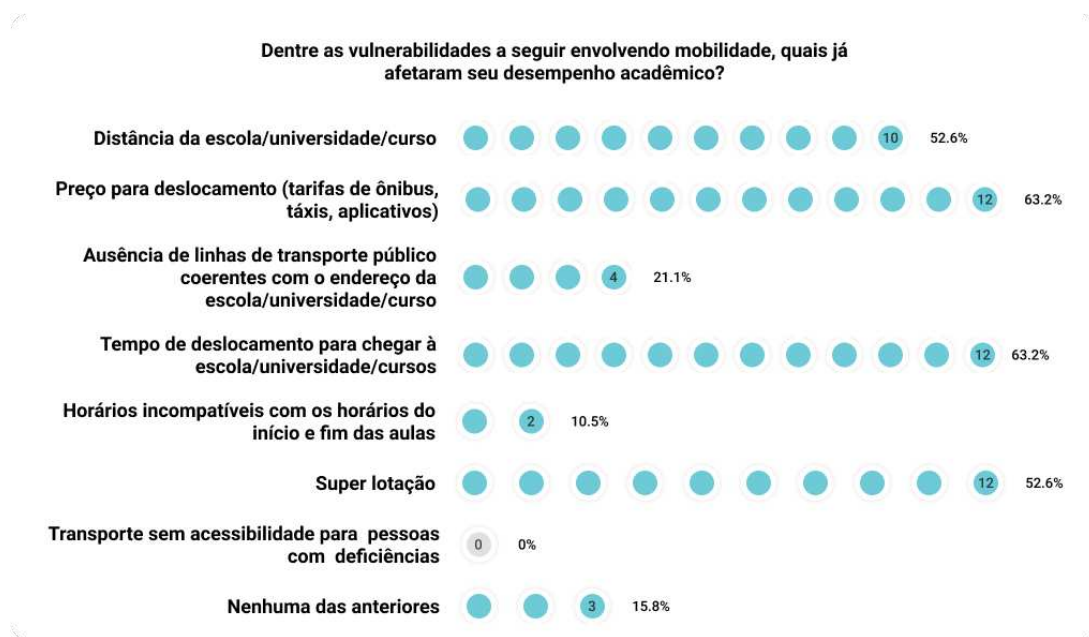


Fonte: Autoria própria

A Figura 8 mostra os resultados para a pergunta sobre as vulnerabilidades percebidas pelos estudantes. Nessa pergunta, os estudantes marcaram as opções que representam vulnerabilidades que acreditam serem capazes de afetar o desempenho acadêmico.

Quatro opções tiveram destaque nessa questão entre os estudantes: Distância entre as residências dos alunos e as instituições de ensino, preços e tarifas para utilizar o transporte público, tempo de deslocamento para chegar nas instituições de ensino e super lotação do transporte público.

Figura 8 – Vulnerabilidades na Mobilidade Urbana - Estudantes



Fonte: Autoria própria

Em geral, as respostas indicam que os estudantes acreditam que as dificuldades na mobilidade urbana têm impactos diretamente negativos ao seu desempenho acadêmico. As principais vulnerabilidades estão relacionadas a distância percorrida e o tempo de deslocamento para chegar às instituições de ensino. Além disso, as demais questões e observações apontam também para o desgaste e o cansaço que o transporte público causa aos estudantes em suas rotinas de estudos.

6.2.2 Especialistas de Educação

A divisão de perguntas do formulário dirigidas aos especialistas de educação contém dezesseis questões. Esta subseção apresenta as respostas para as principais perguntas e as conclusões e observações realizadas.

Figura 9 – Perfil de Especialistas de Educação

Qual a sua relação com o domínio da educação?

8 respostas



Fonte: Autoria própria

Oito especialistas de educação responderam o formulário, a Figura 9 apresenta um gráfico com o perfil desses especialistas. Dentre eles, 25% são professores pesquisadores da área de educação, 12,5% são professores de matemática e formadores de professores do ensino fundamental, 12,5% são professores do ensino médio e superior e 50% professores do ensino fundamental.

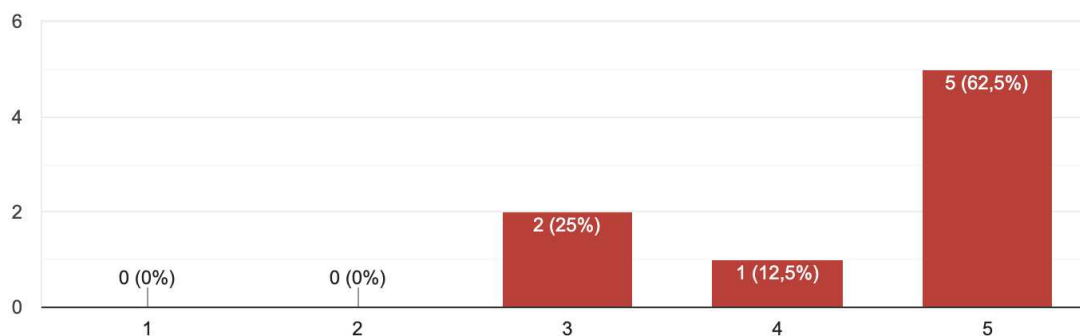
A opinião desses educadores e especialistas condiz com as conclusões realizadas das respostas dos estudantes, já que sete (87,5%) disseram que a longa distância das escolas, universidades e cursos pode gerar uma instabilidade no desempenho acadêmico dos alunos e apenas um (12,5%) afirmou não ter certeza se a distância percorrida pelos estudantes é um fator para a instabilidade acadêmica.

A Figura 10 mostra a o nível de importância dada por estes especialistas à integração dos domínios na solução de problemas da cidade, foi usada uma escala *Likert*, sendo sendo o nível cinco muito importante e o nível um nada importante. Cinco (62,5%) especialistas concederam o nível cinco, apenas um (12,5%) considerou o nível quatro de importância e dois (25%) consideraram como nível três de importância.

Figura 10 – Valor da Visão Integrada dos Domínios

Uma visão integrada desses dois domínios, educação e mobilidade pode trazer benefícios para o desenvolvimento da cidade?

8 respostas

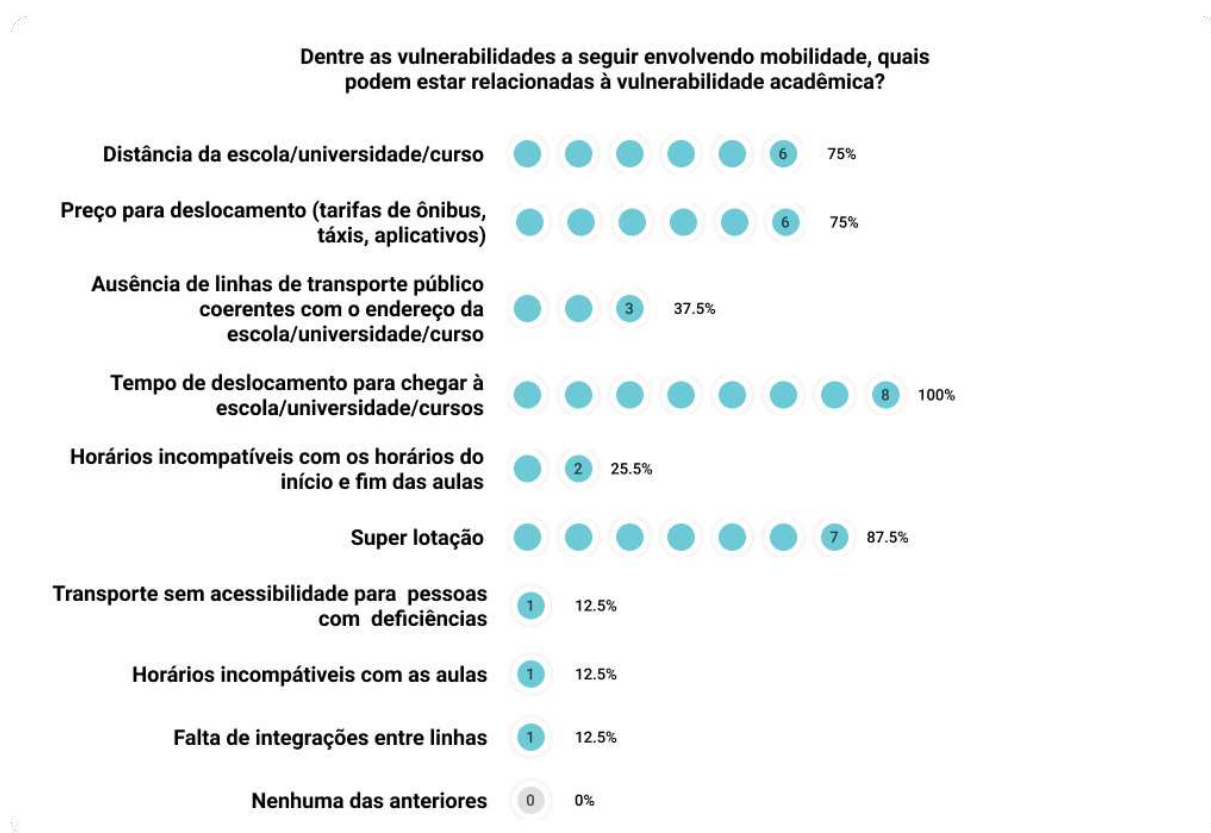


Fonte: Autoria própria

A Figura 11 dispõe os resultados para a pergunta sobre as vulnerabilidades percebidas pelos professores. Nessa pergunta, os educadores especialistas de educação marcaram as opções que representam vulnerabilidades que acreditam serem capazes de afetar o desempenho acadêmico de seus alunos.

Quatro opções tiveram destaque nessa questão entre os estudantes: Distância entre as residências dos alunos e as instituições de ensino, preços e tarifas para utilizar o transporte público, tempo de deslocamento para chegar nas instituições de ensino e super lotação do transporte público. De forma geral, as respostas dadas pelos especialistas de domínio coincidem com as respostas dadas pelos estudantes para a mesma pergunta.

Figura 11 – Vulnerabilidades na Mobilidade Urbana - Educação



Fonte: Autoria própria

Em questões discursivas, os especialistas de educação apontaram algumas consequências que a qualidade da mobilidade urbana pode causar diretamente na educação, como: Evasão dos alunos, taxa de presença dos alunos em aula, nível de atenção dos alunos, atrasos e consequências negativas no desempenho em provas e trabalhos.

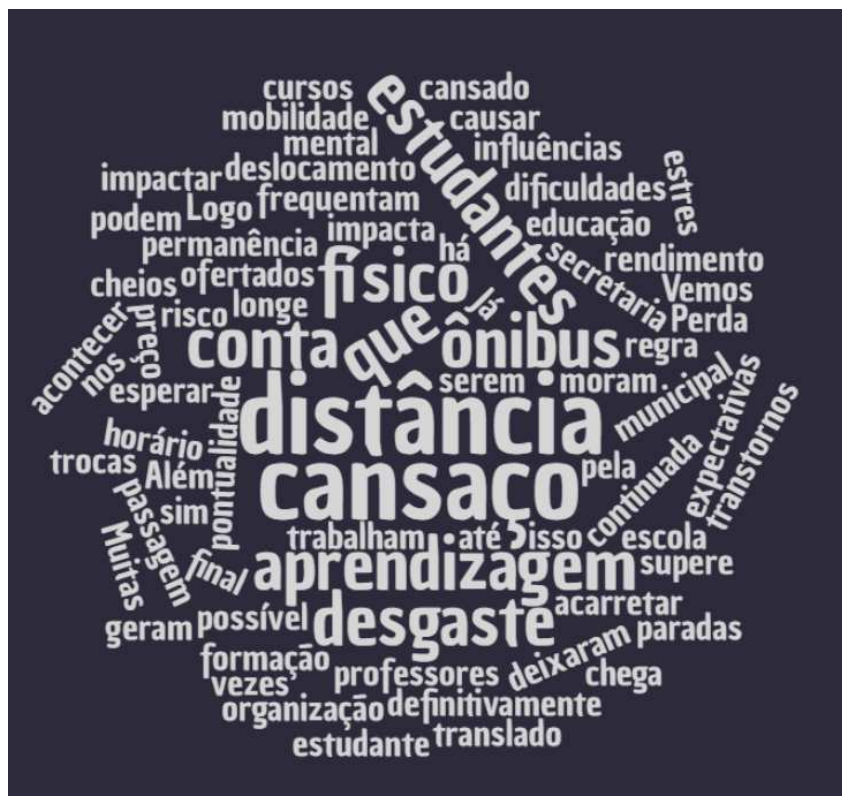
Quando perguntados, em uma questão aberta, sobre suas percepções em relação ao impacto da mobilidade urbana na educação de acordo com suas vivências, os especialistas do domínio de educação apontaram a facilidade de acesso a transporte público e às instituições de ensino como um fator facilitador.

Além disso, outros dois fatores negativos apareceram nas respostas da maioria dos especialistas, diretamente ou indiretamente, são eles: Desgaste físico que o transporte público causa aos estudantes e cansaço em relação ao deslocamento. Esses fatores, segundo eles, influenciam diretamente no nível de atenção dos alunos em sala de aula e na taxa de permanência deles nas instituições de ensino.

A Figura 12 expõe uma nuvem de conceito construída com os termos que mais aparecem nas respostas discursivas dos especialistas do domínio de educação. Estão representadas em fontes maiores os termos que mais apareceram nas respostas, e em fontes menores os termos de menor presença. Os termos de destaque são: Distância, cansaço

físico, aprendizagem, estudantes, desgaste e ônibus.

Figura 12 – Nuvem de Palavras: Educação e Mobilidade Urbana



Fonte: Autoria própria

A última pergunta feita para os educadores e especialistas foi: Quais informações relacionadas à mobilidade urbana você gostaria de ter dos gestores de mobilidade urbana que auxiliariam na gestão da educação em uma cidade?

As principais respostas foram referentes a ter conhecimento sobre as linhas e pontos de acesso à transporte público que estão presentes nas proximidades de determinada instituição de ensino. E sobre entender como as linhas e trajetos foram definidos para oferecer melhor acessibilidade para os estudantes.

6.2.3 Especialistas de Mobilidade Urbana

Como já apontado neste capítulo, pela falta de respostas de especialistas de mobilidade, optou-se por realizar uma entrevista presencial com a especialista de mobilidade urbana selecionada, a professora Tatiana Gadda, docente do departamento acadêmico de arquitetura e urbanismo da UTFPR.

A especialista concordou com as conclusões e observações realizadas pelos outros grupos, de que as dificuldades presentes na mobilidade urbana podem causar influên-

cias diretamente negativas no desempenho acadêmico de estudantes de diferentes níveis educacionais.

A especialista fez algumas observações e indicou estudos importantes sobre o sistema público de transporte da cidade. Baseada no estudo de Duarte et al. (2016), ela apontou que existem 437.059 estudantes do ensino superior no Estado do Paraná, cuja capital é Curitiba. 32% desses estudantes estão concentrados na região metropolitana de Curitiba. Sendo a Universidade Federal do Paraná a principal universidade da região, com cerca de 32.000 alunos, seguida pela Pontifícia Universidade Católica do Paraná e pela Universidade Tecnológica Federal do Paraná, com cerca de 30.000 e 17.000 estudantes, respectivamente.

Segundo Bergman (2017), a cidade de Curitiba, é considerada o berço do sistema *Bus Rapid Transit* (BRT). Um plano diretor aprovado em meados da década de 1960 orientou o desenvolvimento da cidade em uma direção de Desenvolvimento Orientado ao Trânsito (TOD), por zoneamento para altas densidades de desenvolvimento próximas às cinco linhas principais de BRT nos chamados eixos estruturais.

Por fim, Tatiana citou o trabalho de Lima e Fortunato (2017), que apontam os terminais de transporte como elementos privilegiados dos Eixos Estruturais de Curitiba, instituindo-os como elementos de uso efetivo e marcos urbanos emblemáticos do Desenho Urbano da cidade. Segundo as autoras, os terminais são pontos nodais de atração e dispersão de usuários do sistema de transporte público, e promovem o desenvolvimento de atividades comerciais e de serviços nos imóveis do entorno, por obterem caráter de subcentro em alguns pontos.

A última pergunta feita para a professora Tatiana Gadda foi: Quais informações relacionadas à educação você gostaria de ter de forma a auxiliar na atividade de gestão urbana no quesito mobilidade?

As respostas foram iguais às resposta obtidas de especialistas de educação e se referem a ter a informação da quantidade de pontos e terminais de ônibus perto das instituições de ensino e avaliar se os números e médias de pontos de acesso à transporte público são diferentes para cada região da cidade e índices socioeconômicos dos bairros e escolas.

6.3 Questões de Competência

As necessidades de informação levantadas foram: Acesso à informação das linhas de ônibus, quantidade de pontos e terminais de ônibus perto das instituições de ensino e sobre o desempenho acadêmico das escolas. Essas necessidades foram essenciais para definir as questões de competência deste trabalho. Segundo Noy et al. (2001), as questões de competência limitam o escopo da ontologia a ser criada.

Baseado nas respostas dos especialistas de domínio e estudantes, foram desenvolvidas quatro de questões de competência que integram dados e informações de ambos os

domínios, educação e mobilidade urbana. São elas:

1. Quais são as linhas próximas de determinada escola?
2. Quais são os pontos de acesso próximos de determinadas escolas?
3. Qual a relação entre o desempenho acadêmico e o acesso à transporte público?
4. Qual a média de pontos de acesso de mobilidade entre as escolas com um Indicador de Nível Socioeconômico (INSE) específico?

As duas primeiras questões de competência foram definidas com base nas necessidades que os especialistas de educação demonstraram, principalmente nas preocupações relacionadas ao acesso e deslocamentos dos estudantes.

As duas segundas questões foram inspiradas nas necessidades de informações demonstradas pela especialista em mobilidade urbana. Ou seja, demonstrar o impacto da mobilidade nos índices de qualidade de educação e investigar se o acesso a transporte pública é diferente de acordo com a região da cidade, e indicadores socioeconômicos da instituição de ensino.

6.4 Definição de Vocabulário

O objetivo desta etapa é definir o vocabulário da ontologia desenvolvida seguindo as declarações e necessidades dos especialistas de domínio, garantindo que a ontologia respeite o escopo, as limitações e particularidades de cada domínio e suas relações.

Então, com base nas respostas do instrumento de coleta, disponível no Apêndice A, as entrevistas e conversas realizadas com os especialistas de domínio foram definidos os termos que compõem o vocabulário da ontologia.

A ontologia desenvolvida tem o objetivo de representar as principais entidades relacionadas a transporte público, cobrindo termos como pontos de ônibus, linhas de ônibus e terminais de ônibus, e educação, cobrindo o ensino básico, fundamental e superior por meio de métricas de avaliação. Então, a relação entre os dois domínios se dará pela representação do desempenho acadêmico e do acesso ao transporte público. Para a representação do desempenho serão utilizadas métricas, são elas:

1. **Prova Brasil:** Avaliação Nacional do Rendimento Escolar que é realizada a cada dois anos para avaliar o nível de ensino das escolas públicas e privadas sobre Português e Matemática para alunos do quinto e nono ano do ensino fundamental.
2. **ENEM:** O Exame Nacional do Ensino Médio é uma prova anual realizada para avaliar o ensino médio das escolas públicas e privadas do Brasil
3. **ENADE:** O Exame Nacional de Desempenho dos Estudantes é uma prova escrita, aplicada anualmente, usada para avaliação dos cursos de ensino superior brasileiros.

No que tange as métricas relacionadas ao domínio da mobilidade urbana e transporte público, foram selecionadas:

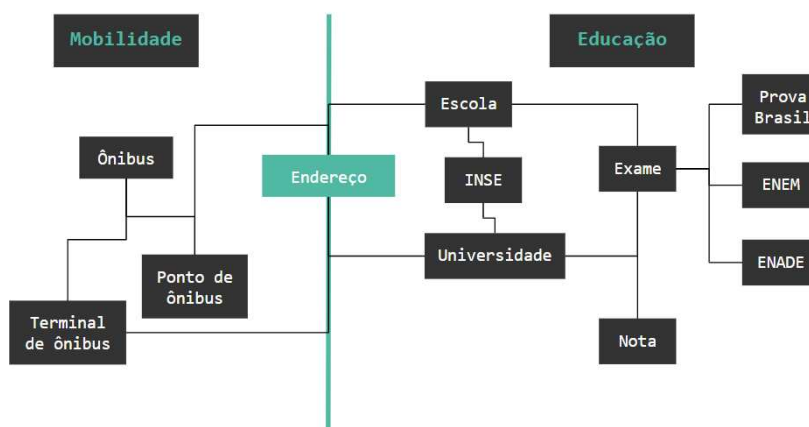
1. **Proximidade de Pontos de Acesso às Instituições de Ensino:** Distância entre as instituições de ensino e os pontos de acesso ao transporte público, pontos de ônibus e terminais de ônibus.
2. **Oferta de Linhas de Ônibus:** Quantidade de linhas de ônibus que podem ser utilizadas pelos estudantes de uma instituição de ensino, dentro de uma determinada distância.

Com o escopo e as métricas definidos, delimitou-se o vocabulário em termos de substantivos e verbos. Os substantivos representam as classes e propriedades presentes no contexto da relação entre os domínios de educação e mobilidade. E os verbos representam as relações entre as entidades presentes no escopo.

1. **Substantivos:** Cidade; Endereço; Logradouro; Bairro; Rua; Prova Educacional; ENEM; ENADE; Prova Brasil; Instituição de Ensino; Ensino Médio; Ensino Superior; Escola; Universidade; Ponto de Ônibus; Terminal de Ônibus; Linha de Ônibus; Sigla; Nome; Ano; Nota; INSE; Código de identificação; Disciplina; Curso de Graduação.
2. **Verbos:** Tem endereço; Tem ônibus; Têm para de ônibus; Tem estatísticas; Está localizado em; Está próximo de; Realiza prova; Possui nome; Possui média em prova.

A Figura 13 mostra os principais conceitos de cada domínio e o que relaciona ambos, o endereço. Do lado direito estão representados os conceitos do domínio de educação e do lado esquerdo os conceitos do domínio de mobilidades. O endereço conecta os domínios pois é a partir dele que é possível saber onde está localizado uma escola, um terminal de ônibus, um ponto de ônibus ou por onde passam as linhas de transporte público da cidade.

Figura 13 – Vocabulário Definido



Fonte: Autoria própria

6.5 Seleção de Ferramentas

Para alcançar os objetivos propostos, foi necessário o suporte de um conjunto de recursos e ferramentas. Inicialmente, deve-se citar os dados que foram utilizados por este trabalho, eles representam os recursos mais essenciais para o desenvolvimento.

O primeiro conjunto de dados se refere aos dados abertos de Curitiba sobre o transporte público armazenados no servidor do projeto EUBra-BigSea³ numa base de dados PostgreSQL5. O acesso a este servidor foi fornecido pela parceria com a professora Nádia P. Kozievitch, docente do departamento acadêmico de informática da Universidade Tecnológica Federal do Paraná.

Segundo os idealizadores do projeto, Alic et al. (2019), EUBra-BIGSEA é uma colaboração destinada a desenvolver serviços de análise de dados baseados na nuvem adaptados principalmente para dados de transporte público.

O servidor do EUBra-BigSea também dispõe de dados sobre a educação no município de Curitiba, mas pretende-se complementar estes dados com o conjunto de microdados disponibilizados em formato aberto pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

Os microdados do INEP⁴ fornecem informações sobre o censo da educação básica e superior de instituições de todo o território nacional e sobre seus desempenhos em exames nacionais, como o Enem e o Enade.

Além desses dois conjuntos de dados, foi utilizado o *Web Service*⁵ da URBS, empresa de economia mista que controla o sistema de transporte público da cidade de

³Disponível em <https://www.eubra-bigsea.eu>

⁴Disponível em <https://portal.inep.gov.br/web/guest/dados>

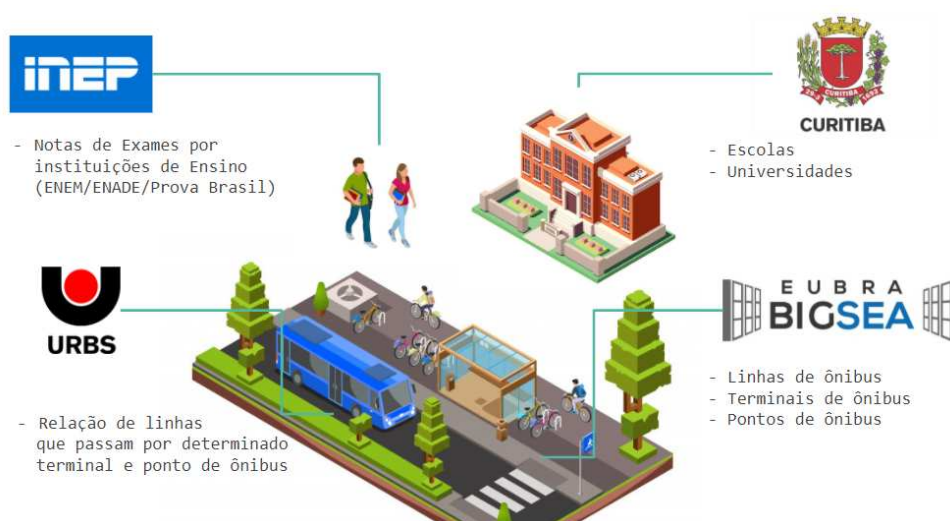
⁵Disponível em <http://transporteservico.urbs.curitiba.pr.gov.br/login.php>

Curitiba. Esse serviço disponibiliza *endpoints* para acesso de informações e dados específicos. Dentre os *endpoints* disponíveis, foi utilizado neste trabalho o *endpoint* para saber por quais pontos de ônibus da cidade determinada linha de transporte passa.

Para poder ter acesso à API da URBS, é necessário enviar um e-mail para relacionamento@urbs.curitiba.pr.gov.br solicitando o acesso por meio da Lei de acesso à informação. A área de Tecnologia da Informação da URBS irá em seguida criar um login e senha para que você tenha acesso ao *Web Service* dos dados públicos do transporte coletivo de Curitiba. O formato dos dados devolvidos pelo serviço é o JSON, mas ele pode facilmente ser convertido para CSV ou outros formatos estruturados de dados.

A Figura 14 mostra as bases de dados abertas utilizados neste trabalho de conclusão de curso e as informações presentes e necessárias em cada uma delas para realizar a conexão dos domínios de educação e mobilidade na cidade de Curitiba.

Figura 14 – Dados abertos utilizados



Fonte: Autoria própria

Para manusear os arquivos das bases foi escolhido a linguagem de programação Python em conjunto da biblioteca Pandas. A biblioteca tem a finalidade de auxiliar na limpeza e trabalho com arquivos CSV's e extração de dados. Ela foi utilizada neste trabalho de conclusão de curso para limpar dados poluídos dentro dos conjuntos disponibilizados em formato aberto pelo INEP e para criar novos CSV's quando necessário.

A biblioteca Pandas disponibiliza funções para leitura e extração de dados de arquivos de dados estruturados. Com esta ferramenta é possível realizar operações complexas sobre o arquivo, como cortes de linhas e colunas.

Com a disponibilidade destes recursos, algumas ferramentas foram utilizadas. A principal delas foi o *software* Protégé, descrito por Noy et al. (2001) como uma ferramenta

gráfica para edição de ontologias e aquisição de conhecimento que possibilita a modelagem conceitual com novas e evolutivas linguagens da Web Semântica. Essa ferramenta foi selecionada por ser um *software open source* que disponibiliza uma vasta gama de *plugins* e por ser bem estabelecido entre a comunidade acadêmica da área de *Web Semântica*.

Além do Protégé, também foi necessário a utilização de um *plugin* específico para o mapeamento de dados de banco relacionais para triplas e grafos, convertendo para a representação do OWL. O Protégé dispõe para essa função um *plugin* chamado Ontop, ele foi utilizado na fase de população da ontologia para auxiliar o processo.

Para manusear e armazenar as triplas geradas foi utilizado o *GraphDB*⁶, ele é um banco de grafos, também conhecido como banco de triplas, responsável por armazenar as triplas geradas pelo *plugin OnTop* e oferecer *endpoints* para realizar consultas complexas, utilizando a linguagem SPARQL.

Quanto à especificação de *hardware*. O desenvolvimento deste projeto se apoia no desenvolvimento da ontologia, que foi desenvolvida em uma máquina com as seguintes configurações: memória RAM de 8,00 GB e processador Intel Core i5-2400 de 3.40GHz.

6.6 Construção da Ontologia

O desenvolvimento desta ontologia, baseou-se nos processos estabelecidos na *Ontology Development 101* de Noy et al. (2001). Como ferramentas de apoio ao desenvolvimento, foram utilizados o *software* Protégé na versão 5.5.0 e um *plugin* chamado *OnTop*⁷ para o mapeamento dos dados em formato relacional para triplas.

A *Ontology Development 101* estabelece os seguintes processos para desenvolvimento de uma ontologia: determinar escopo; considerar reuso; enumerar termos; definir classes; definir propriedades; definir restrições e, criar instâncias. Esses processos, foram agrupados em duas atividades principais: i) especificação da ontologia, ii) aquisição do conhecimento e iii) população da ontologia.

6.6.1 Especificação da Ontologia

A etapa de especificação compreende as seguintes fases da metodologia *Ontology Development 101*: A determinação do escopo da ontologia, reutilização de outras ontologias e determinação de vocabulário. A ontologia desenvolvida tem o objetivo de representar as principais entidades relacionadas a transporte público, cobrindo termos como pontos de ônibus, linhas de ônibus e terminais de ônibus, e educação, cobrindo o ensino básico, fundamental e superior por meio de métricas de avaliação. O vocabulário da ontologia, subetapa da *Ontology Development 101*, já foi definido na seção 6.4, com o título de "Definição de Vocabulário", que representa a etapa seis da metodologia deste trabalho.

⁶Disponível em <http://graphdb.ontotext.com/>

⁷Disponível em <https://github.com/ontop/ontop>

Sobre a determinação do escopo, a ontologia deve ser capaz de descrever as propriedades que estão inseridas em cada uma dessas avaliações e as principais entidades e relações que compõem o contexto de transporte público em ambientes urbanos. A atividade de especificação buscou responder as seguintes perguntas:

a. Qual o propósito da ontologia? A ontologia proposta tem o objetivo conectar as diferentes áreas e domínios do espaço urbano, especificamente mobilidade e educação, e responder a hipótese da relação entre o desempenho acadêmico e o acesso ao transporte público.

b. Qual o escopo da ontologia? Os domínios de Mobilidade Urbana e Educação em cidades inteligentes.

c. Quais fontes de conhecimento e reuso podem ser aplicadas à ontologia? Foi adotada uma ontologia chamada *GeoNames Ontology*⁸ para representar a relação de distância geoespacial entre os objetos. Também foram selecionadas as ontologia *Computer Retrieval of Information on Scientific Projects Thesaurus*⁹ para representar os conceitos de instituições de ensino e a ontologia *Environment Ontology*¹⁰ para representar o conceito de ônibus. Além disso, foram captados conhecimentos de especialistas de domínio das duas áreas.

6.6.2 Aquisição de Conhecimento

A etapa de aquisição de conhecimento compreende as seguintes fases da metodologia *Ontology Development 101*: A definição de termos, classes e propriedades. A atividade de aquisição do conhecimento envolveu processos para conceitualização e formalização da ontologia.

Juntamente com especialistas dos domínios, foram realizadas entrevistas e questionários, e foram definidas as seguintes questões de competências, responsáveis por guiar o desenvolvimento da ontologia:

1. Quais são as linhas próximas de determinada escola?
2. Quais são os pontos de acesso próximos de determinadas escolas?
3. Qual a relação entre o desempenho acadêmico e o acesso à transporte público?
4. Qual a média de pontos de acesso de mobilidade entre as escolas com um INSE específico?

O Quadro 1 mostra a lista de classes definidas na ontologia. Foram definidas quatorze classes, sendo quatro delas reutilizadas de outras ontologias.

⁸Disponível em <http://www.geonames.org/ontology/documentation.html>

⁹Disponível em <http://bioportal.bioontology.org/ontologies/CRISP>

¹⁰Disponível em <http://bioportal.bioontology.org/ontologies/ENVO>

Quadro 1 – Lista de Classes

Axioma Terminológico	Descrição
<i>Educational Statistics</i>	Os fatos e dados numéricos sobre a qualidade da educação de determinado ano.
<i>City</i>	Conglomerado de pessoas que, localizadas em uma área geograficamente delimitada, possuem muitas casas, indústrias, áreas agrícolas.
<i>Address</i>	Dados necessários para localizar uma propriedade (nome da rua, número da casa, apartamento, andar, terreno etc.)
<i>Exam</i>	Avaliação do desempenho acadêmico de uma instituição de ensino.
ENADE	Subclasse da classe <i>Exam</i> , representa o Exame Nacional de Desempenho dos Estudante.
ENEM	Subclasse da classe <i>Exam</i> , representa o Exame Nacional de Desempenho do Ensino Médio.
Prova Brasil	Subclasse da classe <i>Exam</i> , representa a Avaliação Nacional do Rendimento Escolar do ensino fundamental.
<i>School</i>	Estabelecimento destinado ao ensino, público ou privado, representa todas as instituições de ensino.
<i>Secondary School</i>	Subclasse da classe <i>School</i> . Representa uma instituição educacional voltada para o ensino fundamental, fundamental e médio.
<i>University</i>	Subclasse da classe <i>School</i> . Representa uma instituição educacional que abrange o ensino superior, pós-graduação, mestrado, doutorado etc.
<i>Access Point</i>	Passagem física que permite o uso de transporte público.
<i>Bus Stop</i>	Subclasse de <i>Access Point</i> . Representa a parada de um ônibus nas ruas da cidade em seu trajeto.
<i>Bus Station</i>	Subclasse de <i>Access Point</i> . Estruturas onde os ônibus urbanos param para os passageiros embarcarem e / ou desembarcarem.
<i>Bus</i>	Representa qualquer veículo que desempenhe a função de transportar passageiros em uma área urbana.

Fonte: Autoria própria.

A classe *Bus* é original da *Environment Ontology* e da *Computer Retrieval of Information on Scientific Projects Thesaurus* estão sendo utilizadas as classes: *School*, *University* e *Secondary School*.

O Quadro 2 mostra a lista de *Object Properties* definidas na ontologia, juntamente com seus domínios, imagens, relações inversas e descrições. A propriedade intitulada *nearbyFeatures* é proveniente da *GeoNames Ontology*, ela descreve uma relação de proximidade entre dois objetos da ontologia.

A métrica de 500 metros utilizada neste estudo foi definida com base em literaturas da área de arquitetura e urbanismo que referencia uma lei municipal. Esta distância é adotada pelo Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) levando em conta práticas locais e está presente no Artigo 4º da Lei Municipal nº 12.597/2008, onde dispõe sobre a organização do Sistema de Transporte Coletivo Municipal.

Quadro 2 – Lista de Classes

Axioma Terminológico	Domínio	Imagem	Inversa	Descrição
<i>hasAddress</i>	<i>City</i>	<i>Address</i>	<i>isAddressOf</i>	Indica que uma cidade (<i>City</i>) contém um ou mais endereços (<i>Address</i>).
<i>hasBus</i>	<i>Bus Station</i>	<i>Bus</i>	<i>passBy</i>	Indica que um terminal (<i>Bus Station</i>) possui um ou mais ônibus (<i>Bus</i>) que passam por ele.
<i>hasStatistics</i>	<i>Secondary School</i>	<i>Educational Statistics</i>		Uma escola (<i>Secondary School</i>) possui uma estatística educacional (<i>Educational Statistics</i>) para determinado ano.
<i>hasStop</i>	<i>Bus</i>	<i>Bus Stop</i>	<i>isStopOf</i>	Indica que um ônibus (<i>Bus</i>) tem paradas em pontos de ônibus (<i>Bus Stop</i>).
<i>isLocated</i>	<i>Access Point</i> <i>Address</i> <i>School</i>	<i>Address</i>		Indica que algo é localizado por um endereço (<i>Address</i>).
<i>nearbyFeatures</i>	<i>School</i>	<i>Access Point</i>		Indica que a instituição de ensino (<i>Educational Institution</i>) possui, dentro de um raio de 500m, um ponto de acesso (<i>Access Point</i>) próximo.
<i>takesExam</i>	<i>School</i>	<i>Exam</i>	<i>isTakenBy</i>	Indica que uma instituição de ensino (<i>Education Institution</i>) realizou um ou mais exames de avaliação de qualidade de educação (<i>Exam</i>).

Fonte: Autoria própria.

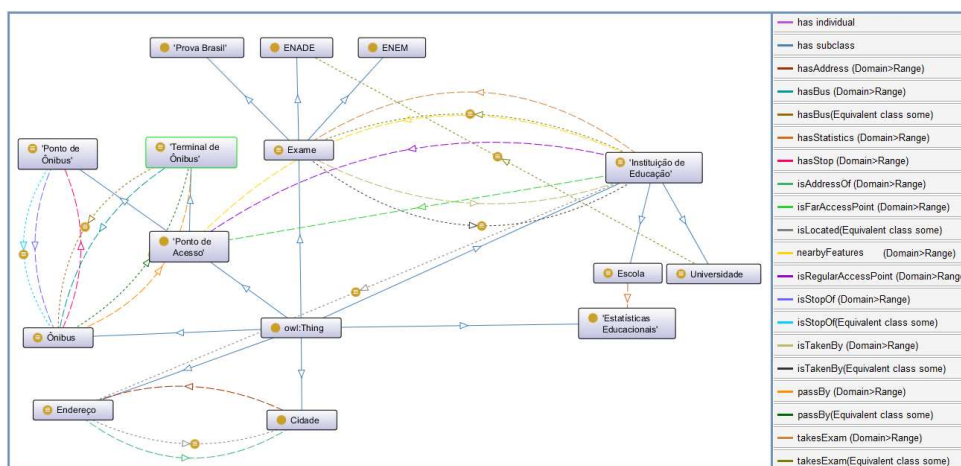
O Quadro 3 mostra a lista de *Data Properties* definidas na ontologia, juntamente com seus domínios e descrições. Essas propriedades foram definidas com base nas informações necessárias para responder as questões de competência definidas e com base na disponibilidade de dados e informações presentes nos dados abertos utilizados para a prova de conceito.

Quadro 3 – Lista de Data Properties

Axioma Terminológico	Domínio	Descrição
<i>hasType</i>	<i>Education Statistics</i>	Define o tipo de determinada escola que realizou o exame: Pública, Particular ou Federal.
<i>hasPermanenceRate</i>	<i>Education Statistics</i>	Taxa de alunos que não abandonaram a escola em um determinado ano.
<i>hasAbandonmentRate</i>	<i>Education Statistics</i>	Taxa de alunos que abandonaram a escola em um determinado ano.
<i>hasApprovalRate</i>	<i>Education Statistics</i>	Taxa de alunos aprovados em uma escola em um determinado ano.
<i>hasDisapprovalRate</i>	<i>Education Statistics</i>	Taxa de alunos reprovados de uma escola em um determinado ano.
<i>hasParticipationRate</i>	<i>Education Statistics</i>	Taxa de participação de uma escola em um determinado ano em um exame.
<i>isExamGrade</i>	Prova Brasil	Nota final da Prova Brasil para determinada escola em uma disciplina.
<i>isExamSubject</i>	Prova Brasil	Disciplina a que se refere a Prova Brasil para determinada escola.
<i>hasProeficientPercentual</i>	Prova Brasil	Percentual de uma turma com notas em nível proeficiente esperado em determinada disciplina.
<i>hasAdvancedPercentual</i>	Prova Brasil	Percentual de uma turma com notas em níveis avançadas em determinada disciplina.
<i>hasAppropriatePercentual</i>	Prova Brasil	Percentual de uma turma com notas em nível médio apropriado em determinada disciplina.
<i>hasBasicPercentual</i>	Prova Brasil	Percentual de uma turma com notas em nível básico em determinada disciplina.
<i>hasInsufficientPercentual</i>	Prova Brasil	Percentual de uma turma com nível insuficiente em determinada disciplina.
	<i>School</i>	
<i>hasCode</i>	Ponto de Acesso Ônibus	Código identificador presente nos Arquivos e bases abertas de dados.
<i>hasCodeIES</i>	<i>School</i>	Código identificador para Institutos de Ensino Superior.
<i>hasINEPCode</i>	<i>School</i>	Código identificador para Instituições de Ensino.
<i>hasDegree</i>	ENADE	Define o curso graduação a que se refere o exame.
<i>hasGeneralGrade</i>	ENADE	Nota média geral do exame.
<i>hasInitials</i>	<i>School</i>	Siglas das instituições de ensino.
<i>hasInse</i>	<i>School</i> <i>Exam</i>	Indicador de Nível Socioeconômico das Escolas de Educação Básica.
<i>hasName</i>	<i>School</i> <i>Bus</i>	Define a palavra ou locução com que se designa o objeto.
<i>hasNeighborhood</i>	<i>Address</i>	Define o nome do bairro ao qual um endereço pertence.
<i>hasNeighborhoodCode</i>	<i>Address</i>	Define o código do bairro ao qual um endereço pertence.
<i>hasCountry</i>	<i>City</i>	Define o país ao qual uma cidade pertence.
<i>hasState</i>	<i>City</i>	Define o estado federativo ao qual uma cidade pertence.
<i>hasYear</i>	<i>Educational Statistics</i> <i>Exam</i>	Ano de determinada estatística ou exame.
<i>essayAverage</i>	ENEM	Nota média da redação no ENEM de determinado ano.
<i>humanSciencesAverage</i>	ENEM	Nota média em ciências humanas no ENEM de determinado ano.
<i>mathAverage</i>	ENEM	Nota média em matemática no ENEM de determinado ano.
<i>naturalSciencesAverage</i>	ENEM	Nota média em ciências naturais no ENEM de determinado ano.
<i>objectiveAverage</i>	ENEM	Nota média em questões objetivas no ENEM de determinado ano.
<i>portugueseAverage</i>	ENEM	Nota média em português no ENEM de determinado ano.
<i>totalAverage</i>	ENEM	Nota média total no ENEM de determinado ano.

A Figura 15 representa as relações entre as os objetos de diferentes classes da ontologia. A Figura possui uma legenda, onde cada *object property* é representada por uma cor diferente.

Figura 15 – Mapa de Relações entre Classes



Fonte: Autoria própria

Os resultados dessa etapa do desenvolvimento, o arquivo *.owl* da ontologia construída e das ontologias utilizadas, estão dispostas abertamente no *Github* através do link: https://github.com/MateusBelizario/ontology_mobed_2020/tree/master/Ontologia.

6.6.3 Tratamento dos Dados

Nesta subseção serão apresentadas com mais detalhes cada base de dados abertos utilizado neste trabalho de conclusão de curso e os tratamentos realizados em cada um deles. Todos os dados utilizados foram coletados em janeiro de 2020, portanto podem não ser os mais atualizados na data de leitura deste trabalho.

A primeira base de dados abertos utilizada foi a de Unidades de Atendimento Ativas de Curitiba, disponível no portal de dados abertos de Curitiba. Esta base contém as unidades de atendimento de uso público, ativas de Curitiba. Ela inclui todas as unidades municipais, bem como unidades estaduais, federais e particulares da cidade. Não foi realizado nenhum tratamento para este conjunto de dados, ele foi utilizado como pivô para realizar filtros em outros conjuntos.

O segundo conjunto utilizado foi o de Microdados do ENADE, disponível no portal de dados abertos do INEP. O conjunto utilizado foi o do ano de 2017, o último disponível em janeiro de 2020, e contém dados das provas do ENADE realizadas no ano de 2017 em todo território nacional.

Inicialmente, foi realizado um corte nos registros dessa base para conter somente as provas realizadas por alunos de universidades da cidade de Curitiba. Depois de realizar

este filtro por localidade, foram filtrados somente as provas de instituições que estivessem contidas na base de dados de Unidades de Atendimento de Curitiba. Este filtro foi utilizado pois um dado importante para este trabalho é a localização da instituição de ensino, e este dado está contido somente na base de Unidades de Atendimento Ativas de Curitiba.

O Apêndice D mostra o código desenvolvido na linguagem *Python* para resolver uma particularidade dessa base. Originalmente a base trás os resultados de provas individuais por alunos das universidades e seria necessário trabalhar num nível de granularidade maior, por universidade, pois todas as outras bases forneciam as médias dos exames de avaliação de qualidade por instituições de ensino. Esse código gera um arquivo CSV com os códigos da universidade e nota média no ENADE para cada curso da instituição de ensino.

A terceira base utilizada foi a do ENEM por escola, também disponível no portal de dados abertos do INEP. Este arquivo contém as notas médias no ENEM por escola entre os anos de 2005 à 2015, a última atualização dessa base foi em 13 de agosto de 2019. Assim como para a base do ENADE, foi realizado um corte nos registros dessa base para conter somente as médias de escolas da cidade de Curitiba.

Depois do corte inicial, foi desenvolvido um código em *Python*, disposto no Apêndice C, para realizar limpezas mais específicas na base. Este código gera um CSV com as médias do ENEM para o ano de 2015, que tenham um INSE para este ano e ainda trata o valor deste campo. O campo com o valor do INSE deve ser tratado pois as tabelas representam os valores de formas diferentes, algumas utilizam a *label* do nível, como "Muito Alta", outras utilizam a *string* com o nível explícito, "Nível 7" e outras somente o número inteiro do nível.

O valor do INSE de 2015¹¹ varia de 1 (Muito Baixo) à 6 (Muito Alto). Para o cálculo de índice, as questões utilizadas dizem respeito à renda familiar, à posse de bens e à contratação de serviços de empregados domésticos pela família dos estudantes, além do nível de escolaridade de seus pais ou responsáveis.

A terceira base de dados foi o da Prova Brasil de 2017¹². Neste caso, foram tratados quatro arquivos CSVs, pois cada uma das séries, 5º ano e 9º ano, realizam duas provas, português e matemática. Inicialmente, como as outras bases, foram realizados cortes para se trabalhar somente com escolas de Curitiba. O Apêndice B mostra o código *Python* desenvolvido para limpar registros com valores nulos e tratar o valor do INSE para cada um destas bases.

Todos os conjuntos de dados utilizados sobre o domínio de mobilidade foram retirados do servidor do projeto EUBra-BIGSEA, com exceção de dois. Deste servidor foram exportadas três tabelas de dados relacionais: a relação de linhas de ônibus de Curitiba, terminais de ônibus de Curitiba e pontos de ônibus de Curitiba.

¹¹Disponível em http://download.inep.gov.br/informacoes_e_statisticas/indicadores_educacionais/2015/nota_tecnica/nota_tecnica_inep_inse_2015.pdf

¹²Disponível em <https://www.qedu.org.br/>

Além deste servidor, foi utilizado o *Web Service* da URBS, para realizar o mapeamento das linhas de ônibus que passam por cada ponto de ônibus da cidade de Curitiba e gerar um CSV com esses dados. O código *Python* utilizado para fazer as requisições ao *endpoint* necessário está disposto no Apêndice I.

Além desses arquivos CSVs gerados, há mais quatro que foram feitos para representar a relação das distâncias entre as instituições de ensino e os pontos de acesso à transporte público, pontos e terminais de ônibus.

O Apêndice E dispõe o código usado para construir o CSV com a relação de proximidade entre as escolas e terminais de ônibus usando a latitude e longitude da base de dados de Unidades de Atendimento Ativas de Curitiba para as instituições de ensino e a latitude e longitude dos pontos de acesso à transporte público dispostas na base de dados do projeto EUBra-BIGSEA.

Seguindo a mesma lógica, o Apêndice F apresenta o código para construir o CSV com a mesma relação para as escolas e pontos de ônibus, o Apêndice G para montar a relação entre universidades e terminais de ônibus e, por último, o Apêndice H dispõe o código *Python* para gerar o CSV com as relações de distância entre as universidades e os pontos de ônibus de Curitiba.

O último arquivo CSV foi gerado manualmente para fazer a relação das linhas de ônibus que passam por cada terminal da cidade. Para montar este arquivo, foram pesquisados os terminais da cidade¹³ e verificadas as linhas que passam por eles, utilizando o site da URBS.

Depois de ter gerados todos os CSV com os dados necessários tratados e limpos, os dados foram importados para tabelas relacionais em uma base de dados PostgreSQL5. A base de dados construída para armazenar os dados tratados possui quatorze tabelas, descritas na Tabela 4

Tabela 4 – Lista de Tabelas Relacionais

Tabela	Descrição
<i>enade</i>	Dados do ENADE para o ano de 2017 das universidades de Curitiba
<i>enem</i>	Dados do ENEM para o ano de 2015 das escolas de Curitiba
<i>prova_brasil_2017</i>	Dados da Prova Brasil para o ano de 2017 das escolas de Curitiba
<i>escolas_curitiba</i>	Tabela com a relação das escolas de Curitiba
<i>universidade_curitiba</i>	Tabela com a relação das universidades de Curitiba
<i>linhas_onibus</i>	Tabela com a relação das linhas de ônibus de Curitiba
<i>terminal_onibus</i>	Tabela com a relação dos terminais de ônibus de Curitiba
<i>pontos_curitiba</i>	Tabela com a relação dos pontos de ônibus de Curitiba
<i>terminal_onibus_pontos_linhas</i>	Tabela com a relação dos ônibus que passam por cada terminal
<i>escola_terminal</i>	Tabela com a relação das linhas que passam por cada ponto
<i>escola_ponto</i>	Tabela com a relação de proximidade entre escolas e terminais
<i>universidade_terminal</i>	Tabela com a relação de proximidade entre escolas e pontos
<i>universidade_ponto</i>	Tabela com a relação de proximidade entre universidades e terminais

¹³Disponível em <https://www.urbs.curitiba.pr.gov.br/comunidade/terminais-de-onibus>

Os arquivos gerados sobre o domínio de mobilidade foram disponibilizados abertamente e podem ser acessados através do *Github*, pelo link: https://github.com/MateusBelizario/ontology_mobed_2020/tree/master/CSV_Mobilidade. Eles foram disponibilizados por se tratarem de mapeamentos não disponibilizados pela URBS ou qualquer outra base de dados abertos do município de Curitiba.

6.6.4 Conexão dos Dados dos Domínios

A etapa de conexão dos dados dos domínios compreende a seguinte fase da metodologia *Ontology Development 101*: criação das instâncias. Com as tabelas contendo os dados tratados e filtrados, foram realizados quinze mapeamentos com o objetivo de transformar os dados em formato relacional em triplas com dados conectados semanticamente .

Os Algoritmos a seguir mostram os mapeamentos realizados na ferramenta *Protégé*, juntamente com o *plugin OnTop*. Os termos entre colchetes representam colunas das tabelas relacionais construídas, demonstradas na subseção anterior.

O Algoritmo 1 representa o mapeamento das instâncias que são endereços de escolas, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela `escolas_curitiba`.

```
1 :endereco_{CD_EQUI} a :Address;
2 :hasName {NM_RUA};
3 :hasType {DS_DEP_ADMINISTRATIVA};
4 :hasNeighborhood {NM_BAIRRO}.
```

Algoritmo 1 – Mapeamento de Endereços de Escolas

O Algoritmo 2 representa o mapeamento das instâncias das escolas de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela `escolas_curitiba`. A última linha deste mapeamento é a definição de uma *Object Property* que é a relação com seu endereço definido no mapeamento anterior.

```
1 :escola_{CD_EQUI} a csp:0988-5661;
2 :hasName {NM_ABREV_EQUI};
3 :hasCode {CD_EQUI};
4 :hasINEPCode {CD_LOCAL};
5 :hasType {DS_DEP_ADMINISTRATIVA};
6 :isLocated :endereco_{CD_EQUI} .
```

Algoritmo 2 – Mapeamento de Escolas

O Algoritmo 3 representa o mapeamento das instâncias que são endereços de universidade, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com

dados da tabela universidades_curitiba.

```

1 :endereco_{COD_IES} a :Address;
2 :hasName {NM_RUA};
3 :hasType {DS_DEP_ADMINISTRATIVA};
4 :hasNeighborhood {NM_BAIRRO} .

```

Algoritmo 3 – Mapeamento de Endereços de Universidades

O Algoritmo 4 representa o mapeamento das instâncias das escolas de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela universidades_curitiba. A última linha deste mapeamento é a definição de uma *Object Property* que é a relação com seu endereço definido no mapeamento anterior.

```

1 :universidade_{COD_IES} a csp:0989-1313;
2 :hasName {NM_ABREV_EQUI};
3 :hasCode {CD_EQUI};
4 :hasCodeIES {COD_IES};
5 :hasInitials {NM_ABREV_EQUI} ;
6 :isLocated :endereco_{COD_IES} .

```

Algoritmo 4 – Mapeamento de Universidades

O Algoritmo 5 representa o mapeamento das instâncias das médias do enade das universidade de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela enade. A última linha deste mapeamento é a definição de uma *Object Property* que é a relação com a universidade que possui essa nota média do enade para um curso específico.

```

1 :enade_{cod_ies}_{co_curso} a :National_Student_Performance_Exam ;
2 :hasYear {ano} ;
3 :hasDegree {co_curso} ;
4 :hasGeneralGrade {nt_geral} ;
5 :isTakenBy :universidade_{cod_ies} .

```

Algoritmo 5 – Mapeamento ENADE

O Algoritmo 6 representa o mapeamento das instâncias das médias da Prova Brasil das escolas de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela prova_brasil_2017. A última linha deste mapeamento é a definição de uma *Object Property* que é a relação com a escola que possui essa nota média da Prova Brasil.

```

1 :prova_brasil_{Codigo_INEP}_{Disciplina}_{Serie} a :National_School_Performance_Exam ;
2 :isExamSubject {Disciplina} ;

```

```

3 :isExamGrade {Serie} ;
4 :hasInsufficientPercentual {Percentual_Insuficiente} ;
5 :hasBasicPercentual {Percentual_Basico} ;
6 :hasProefficientPercentual {Percentual_Proficiente} ;
7 :hasAdvancedPercentual {Percentual_Avancado} ;
8 :hasAppropriatePercentual {Percentual_Aprendizado_Adequado} ;
9 :hasINSE {NSE_classe} ;
10 :hasParticipationRate {Taxa_Participacao} ;
11 :isTakenBy :escola_{CD_EQUI} .

```

Algoritmo 6 – Mapeamento Prova Brasil

O Algoritmo 7 representa o mapeamento das instâncias das médias do ENEM das escolas de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela enem.

A última linha deste mapeamento é a definição de uma *Object Property* que é a relação com a escola que possui essa nota média do enem. E a penúltima linha representa a relação com as estatísticas educacionais daquele mesmo ano para essa escola.

```

1 :enem_{CO_ESCOLA_EDUCACENSO}_{NU_ANO} a :National_High_School_Exam ;
2 :hasYear {NU_ANO} ;
3 :hasParticipationRate {NU_TAXA_PARTICIPACAO} ;
4 :naturalSciencesAverage {NU_MEDIA_CN} ;
5 :humanSciencesAverage {NU_MEDIA_CH} ;
6 :portugueseAverage {NU_MEDIA_LP} ;
7 :mathAverage {NU_MEDIA_MT} ;
8 :essayAverage {NU_MEDIA_RED} ;
9 :objectiveAverage {NU_MEDIA_OBJ} ;
10 :totalAverage {NU_MEDIA_TOT} ;
11 :hasStatistics :statistics_{NU_ANO}_{CO_ESCOLA_EDUCACENSO} ;
12 :isTakenBy :escola_{CD_EQUI} .

```

Algoritmo 7 – Mapeamento Prova ENEM

O Algoritmo 8 representa o mapeamento das estatísticas educacionais feitas pelo ENEM para as escolas de Curitiba, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela enem.

```

1 :statistics_{NU_ANO}_{CO_ESCOLA_EDUCACENSO} a :Educational_Statistics ;
2 :hasType {TP_DEPENDENCIA_ADM_ESCOLA} ;
3 :hasAbandonmentRate {NU_TAXA_ABANDONO} ;
4 :hasApprovalRate {NU_TAXA_APROVACAO} ;
5 :hasDisapprovalRate {NU_TAXA_REPROVACAO} ;
6 :hasPermanenceRate {NU_TAXA_PERMANENCIA} ;
7 :hasYear {NU_ANO} ;
8 :hasINSE {INSE} .

```

Algoritmo 8 – Mapeamento Estatísticas Educacionais

O Algoritmo 9 representa o mapeamento das linhas de ônibus, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são

Data Properties que a instância possui, populada com dados da tabela `linhas_onibus`.

```
1 :linha_{cod_linha} a envo:ENVO_01000611 ;
2 :hasCode {cod_linha} ;
3 :hasName {nome_linha} .
```

Algoritmo 9 – Mapeamento de Linhas de Ônibus

O Algoritmo 10 representa o mapeamento dos terminais de ônibus, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela `terminal_onibus`.

```
1 :terminal_{id} a :Bus_Station ;
2 :hasName {nome} ;
3 :hasCode {id} .
```

Algoritmo 10 – Mapeamento de Terminal de Ônibus

O Algoritmo 11 representa o mapeamento dos pontos de ônibus, a primeira linha é a identificação da instância e a classe que ela assume na ontologia. As linhas seguintes são *Data Properties* que a instância possui, populada com dados da tabela `pontos_curitiba`.

```
1 :ponto_{NUM} a :Bus_Stop ;
2 :hasCode {NUM} ;
3 :hasName {NOME} .
```

Algoritmo 11 – Mapeamento de Ponto de Ônibus

O Algoritmo 12 é o mapeamento das linhas que passam por determinado terminal de ônibus. A propriedade `:hasBus` define essa relação entre instâncias da classe *Bus Station* e *Bus*. Para popular essa relação foram utilizados os dados da tabela `terminal_onibus`.

```
1 :terminal_{terminal} :hasBus :linha_{onibus}.
```

Algoritmo 12 – Mapeamento de Linha de Ônibus-Terminal de Ônibus

O Algoritmo 13 é o mapeamento das linhas que passam por determinado ponto de ônibus. A propriedade `:hasBus` define essa relação entre instâncias da classe *Bus Stop* e *Bus*. Para popular essa relação foram utilizados os dados da tabela `pontos_linhas`.

```
1 :ponto_{NUM} :hasBus :linha_{CODE}.
```

Algoritmo 13 – Mapeamento de Linha de Ônibus-Ponto de Ônibus

O Algoritmo 14 é o mapeamento das escolas que estão próximas de determinados pontos e estações de ônibus. A propriedade `geo:nearbyFeature` define essa relação entre instâncias da classe *Secondary School* e *Access Point*. Para popular essa relação foram utilizados os dados das tabelas `escola_ponto` e `escola_terminal`.

```
1 :escola_{id} geo:nearbyFeature :terminal_{id}
2 :escola_{id} geo:nearbyFeature :ponto_{id}
```

Algoritmo 14 – Mapeamento Distância Escola-Terminal/Ponto

O Algoritmo 15 é o mapeamento das universidades que estão próximas de determinados pontos e estações de ônibus. A propriedade *geo:nearbyFeature* define essa relação entre instâncias da classe *University* e *Access Point*. Para popular essa relação foram utilizados os dados das tabelas *universidade_ponto* e *universidade_terminal*.

```
1 :universidade_{id} geo:nearbyFeature :terminal_{id}
2 :universidade_{id} geo:nearbyFeature :ponto_{id}
```

Algoritmo 15 – Mapeamento Distância Universidade-Terminal/Ponto

Como resultado da população da ontologia, obteve-se um arquivo em formato *textturtle* (.ttl), com as triplas geradas pelo *plugin OnTop*. No geral, foram criadas 114.605 triplas, e o arquivo com os dados conectados está disposto abertamente no *Github* e pode ser acessado através do *link*: https://github.com/MateusBelizario/ontology_mobed_2020/tree/master/Dados_Conectados.

7 RESULTADOS

Neste capítulo serão mostrados os resultados obtidos com o desenvolvimento deste trabalho e serão dispostas as *queries SPARQL* realizadas no banco de grafos *GraphDB* que respondem as questões de competências definidas. A Figura 16 mostra a *query* que responde a primeira questão de competência Q1: *Quais são as linhas próximas de determinada escola?*

Figura 16 – Questão de Competência 1

```

1 prefix mbed:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 select DISTINCT ?busName where {
5 {
6 {
7   ?school gn:nearbyFeatures ?accessPoint.
8   ?accessPoint mbed:hasBus ?bus.
9   ?bus mbed:hasName ?busName.
10  ?school mbed:hasCode 2246.
11 }
12 }
13 UNION
14 {
15 {
16   ?school gn:nearbyFeatures ?accessPoint.
17   ?accessPoint mbed:isStopOf ?bus.
18   ?bus mbed:hasName ?busName.
19   ?school mbed:hasCode 2246.
20 }
21 }
22 }

```

Fonte: Autoria própria

Na *query* apresentada é feita uma união lógica entre as linhas de ônibus que passam perto da escola, por meio das paradas de ônibus e pelos terminais de ônibus próximos para a escola DOM BOSCO (Unidade Batel), com o código de identificação 2246, sendo Batel um Bairro central de Curitiba. E a Figura 17 mostra seu resultado, a lista com o nome das linhas de ônibus que passam dentro de um raio de 500 metros da escola.

Figura 17 – Resultado: Questão de Competência 1

	busName
1	CAMP.SIQ./BATEL
2	JD.SOCIAL/BATEL
3	RUA XV / BARIGUI
4	TRAMONTINA
5	V. SANDRA
6	AHÚ/LOS ANGELES

Fonte: Autoria própria

A Figura 18 mostra a *query* que responde a segunda questão de competência Q2: *Quais são os pontos de acesso próximos de determinada escola?* Nela são selecionados os pontos de acesso próximos à escola DOM BOSCO, unidade Batel, com o código de identificação 2246. E a Figura 17 mostra seu resultado, a lista que relaciona a escola e o nome dos pontos de acesso próximos, que são o próprio endereço das paradas ou do terminal, se houvesse algum.

Figura 18 – Questão de Competência 2

```

1 prefix mobed:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 select ?schoolName ?accessName where {
5   ?school gn:nearbyFeatures ?accessPoint.
6   ?school mobed:hasCode 2246 .
7   ?school mobed:hasName ?schoolName.
8   ?accessPoint mobed:hasName ?accessName.
9 }

```

Fonte: Autoria própria

Como pode-se ver na Figura 19, há três pontos de ônibus localizados dentro de um raio de 500m dessa escola, o endereço deles está indicado na coluna *accessName*. A ontologia ainda permite pesquisa de quais são os ônibus que passam por esses pontos, caso seja de interesse dos gestores e cidadãos, usando a propriedade *isStopOf* definida.

Figura 19 – Resultado: Questão de Competência 2

	schoolName	accessName
1	Dom Bosco - Batel	Rua Gonçalves Dias, 509 - Batel
2	Dom Bosco - Batel	Rua Bispo Dom José, 130 - Batel
3	Dom Bosco - Batel	Av. Sete de Setembro, 6001 - Batel

Fonte: Autoria própria

A Figura 20 mostra a *query* para um caso de avaliação sobre o ENEM, que tenta responder a terceira questão de competência Q3: *Qual a relação entre o desempenho acadêmico e o acesso ao transporte público?* Nessa *query* é realizada a seleção das diferentes sessões da prova do ENEM e a quantidade de pontos de acesso para cada escola, que podem ser ordenadas por qualquer disciplina.

Figura 20 – Questão de Competência 3

```

1 prefix moped:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 select ?name (count(?accessPoint) AS ?accessPointNumber) ?natural ?human ?portuguese ?essay
5 WHERE {
6     ?exam moped:isTakenBy ?institution.
7     ?institution moped:hasName ?name.
8     ?exam a moped:National_High_School_Exam.
9     ?institution gn:nearbyFeatures ?accessPoint.
10    ?exam moped:naturalSciencesAverage ?natural.
11    ?exam moped:humanSciencesAverage ?human.
12    ?exam moped:portugueseAverage ?portuguese.
13    ?exam moped:mathAverage ?math.
14    ?exam moped:essayAverage ?essay
15 }
16 GROUP BY ?institution ?natural ?human ?portuguese ?essay ?name
17 ORDER BY ?natural

```

Fonte: Autoria própria

E a Figura 21 mostra seu resultado, onde as escolas estão sendo ordenadas pela nota da disciplina de ciências naturais. O resultado mostra somente as cinco escolas com as menores notas em ciências naturais, mas originalmente a *query* traz todas as escolas ordenadas. Os resultados da *query* não mostram de forma clara a relação direta entre os domínios, porém fornece os dados necessários para realizar investigações mais completas utilizando análises estatísticas.

Figura 21 – Resultado: Questão de Competência 3

	name	accessPointNumber	natural	human	portuguese	essay
1	São Sebastião	"2"	"4,5306E2"	"5,466E2"	"4,8111E2"	"5,6714E2"
2	João Mazzarotto	"3"	"4,6214E2"	"5,3317E2"	"5,0363E2"	"5,1931E2"
3	Paulo Leminski	"3"	"4,6441E2"	"5,4832E2"	"5,0708E2"	"5,3774E2"
4	Anibal Khury Neto	"7"	"4,6477E2"	"5,5499E2"	"4,9744E2"	"5,0549E2"
5	La Salle	"2"	"4,6509E2"	"5,5545E2"	"4,9985E2"	"5,2272E2"

...

Fonte: Autoria própria

A Figura 22 mostra a *query* que responde a última questão de competência Q4: *Qual a média de pontos de acesso de mobilidade dentre as escolas com um INSE específico?* No exemplo, está sendo selecionada a média de pontos de acesso para as escolas com o INSE pertencente ao grupo 5 ¹.

Figura 22 – Questão de Competência 4

```

1 prefix mobed:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 SELECT (AVG(?accessNumber) AS ?avg_inse)
5 {
6     SELECT ?institution (count(DISTINCT ?accessPoint) as ?accessNumber)
7     WHERE {
8         ?institution gn:nearbyFeatures ?accessPoint.
9         ?exam mobed:isTakenBy ?institution.
10        ?exam mobed:hasStatistics ?statistics.
11        ?statistics mobed:hasINSE 5
12    }
13 GROUP BY ?institution

```

Fonte: Autoria própria

A Figura 23 mostra o resultado para escolas com o INSE igual a 5. Na cidade há escolas com o INSE variando de 4 à 6, onde a médias de pontos de acesso (terminais e pontos de ônibus) próximos às escolas é maior para instituições de ensino com o INSE 4, atingindo cerca de 6.3 pontos de acesso a transporte público. Escolas com INSE 5 e 6 tiveram média semelhantes, 4.03 e 4.75 respectivamente.

¹O INSE pode variar de 1 a 6, sendo 6 o grupo de maior nível e que indica que os alunos em geral possuem em suas casas um quantitativo alto de bens elementares. Disponível em <http://portal.inep.gov.br/web/guest/indicadores-educacionais>

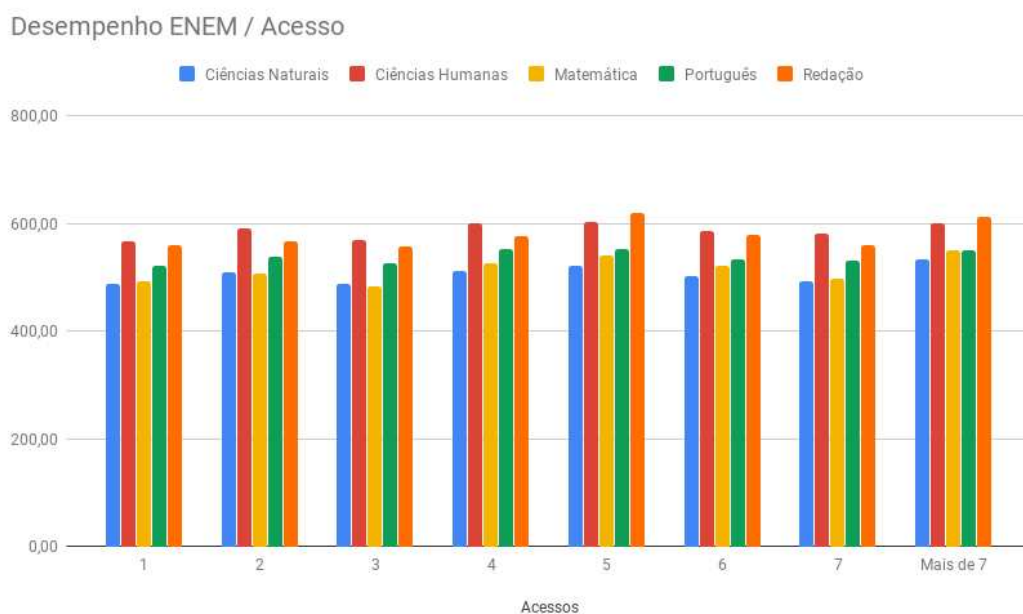
Figura 23 – Resultado: Questão de Competência 4

	avg_inse
1	"4.038461538461538461538462"::xsd:decimal

Fonte: Autoria própria

A Figura 24 mostra um gráfico construído com os resultados obtidos da consulta realizada na Figura 20. Neste gráfico foram plotadas as médias para cada disciplina presente na prova do ENEM agrupadas pelo número de pontos de acesso próximos às instituições de ensino. De forma geral, as médias tiveram uma leve melhora até as instituições com cinco pontos de acesso.

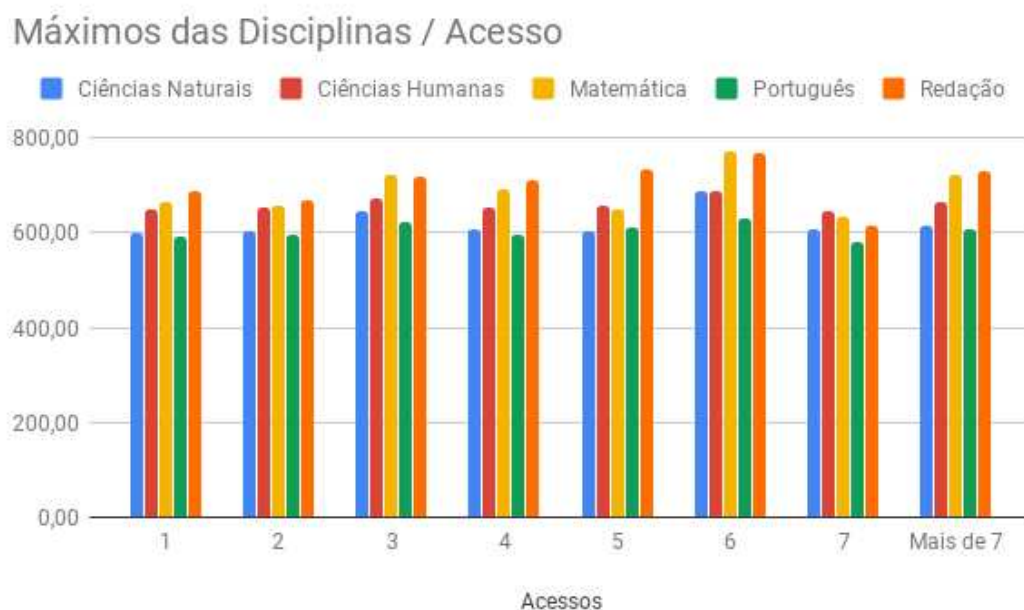
Figura 24 – Desempenho Médio no ENEM por Pontos de Acesso Próximos



Fonte: Autoria própria

Buscando dados que pudessem reforçar a ideia do impacto da mobilidade urbana na educação, decidiu-se explorar os máximos das notas para cada disciplina categorizadas pelo número de pontos de acesso próximos às instituições de ensino. O resultado está presente na Figura 25.

Figura 25 – Máximo Desempenho no ENEM por Pontos de Acesso Próximos



Fonte: Autoria própria

Percebeu-se que as maiores notas de cada disciplinas do ENEM, com poucas exceções, melhoraram conforme o número de pontos de acesso próximos às instituições de ensino aumentaram até haver seis pontos de acesso a transporte público. Esses dados explicitam a relação entre os domínios e podem indicar que o acesso a mobilidade urbana pode ser um fator que contribui para a melhora do desempenho acadêmico.

Para explorar mais a conexão dos dados, decidiu-se realizar mais duas consultas: A quantidade de acessos a transporte público próximos de instituições de ensino por bairro de Curitiba e a média de pontos de acesso a transporte público de instituições de ensino por tipo de instituição (particular, público estadual ou publico municipal).

A Figura 26 mostra a *query* para realizar a consulta sobre a quantidade de acessos a transporte público próximos de instituições de ensino por bairro de Curitiba.

Figura 26 – Consulta de Acessos por Bairro de Curitiba

```

1 prefix moped:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 SELECT ?neighborhood (COUNT(DISTINCT ?accessPoint) AS ?accessNumber)
5 WHERE {
6     ?institution gn:nearbyFeatures ?accessPoint.
7     ?institution moped:isLocated ?address.
8     ?address moped:hasNeighborhood ?neighborhood.
9 }
10 GROUP BY ?neighborhood
11 ORDER BY DESC(?accessNumber)

```

Fonte: Autoria própria

Como observa-se na Figura 27 mostra o resultado para a *query* acima, os cinco bairros com mais pontos de acesso próximos de instituições de ensino são respectivamente: Cidade Industrial de Curitiba (CIC), o maior bairro de Curitiba, Sítio Cercado, Centro, Uberaba e Portão.

Figura 27 – Quantidade de Acessos a Transporte Público por Bairro

	neighborhood	accessNumber
1	Cidade Industrial de Curitiba	"201"
2	Sítio Cercado	"103"
3	Centro	"94"
4	Uberaba	"85"
5	Portão	"76"

Fonte: Autoria própria

A Figura 28 mostra a *query* para realizar consulta sobre a média de pontos de acesso a transporte público de instituições de ensino por tipo de instituição (particular, público estadual ou público municipal).

Figura 28 – Consulta de Acessos por Tipo de Instituição

```
1 prefix mobed:<http://www.semanticweb.org/mateus/ontologies/2019/9/mobility_&_education#>
2 PREFIX gn: <https://www.geonames.org/ontology#>
3
4 SELECT (AVG(?accessNumber) AS ?avg_tipo)
5 {
6   SELECT ?institution (count(DISTINCT ?accessPoint) as ?accessNumber)
7   WHERE {
8     ?institution gn:nearbyFeatures ?accessPoint.
9     ?institution mobed:hasType "Público Estadual"
10  }
11 GROUP BY ?institution
12 }
```

Fonte: Autoria própria

Os resultados mostraram que a média de acessos por tipo da instituição são muito semelhantes. Instituições particulares tem em média 4.63 pontos de acesso a transporte público dentro de um raio de 500m, instituições públicas municipais tem em média 4.53 pontos de acesso e instituições públicas estaduais possuem em média 4.32 pontos de acesso a transporte público próximas.

Pela integração dos dados abertos sobre os domínios, foi possível obter as linhas de transporte público que estão disponíveis dentro de um raio de 500 metros da escola, descobrir os pontos e terminais de ônibus próximos à escola, realizar uma *query* para descobrir as notas médias em avaliações de desempenho acadêmico para cada escola e a média de pontos de acesso a transporte público de acordo com o INSE da instituição de ensino.

A principal dificuldade dessa fase foi responder a terceira questão de competência, pois a relação entre os domínios não é direta. Há diversos fatores presentes na definição do desempenho acadêmico de estudantes de uma instituição de ensino, e somente a quantidade de pontos de acesso e disponibilidade de opções de transporte público não sejam suficientes para esclarecer essa relação. Entretanto, com os dados obtidos pela *query*, disponível na Figura 20, foi possível realizar análises mais completas desse cenário.

8 CONCLUSÃO

Neste trabalho de conclusão foram apresentados problemas sobre a integração semântica de dados em cidades inteligentes, especificamente sobre o caso de mobilidade urbana e educação no município de Curitiba. Este trabalho propôs a integração semântica de dois domínios do espaço urbano, especificamente mobilidade e educação, abrangendo conceitos e soluções *Linked Open Data*. Foram delimitados quatro objetivos específicos para resolver este problema.

O primeiro deles se tratava de conhecer os dois domínios e identificar possíveis benefícios e pontos de integração. Este objetivo foi tratado com o formulários, entrevistas e questionários realizados com os especialistas de domínios e estudantes.

Os estudantes acreditam que as dificuldades na mobilidade urbana têm impactos diretamente negativos ao seu desempenho acadêmico. As principais vulnerabilidades estão relacionadas á distância percorrida e o tempo de deslocamento para chegar as instituições de ensino. Os especialistas da área de educação gostariam ter conhecimento sobre as linhas e pontos de acesso à transporte público que estão presentes nas proximidades de determinada instituição de ensino. E a especialista de mobilidade entrevistada gostaria de saber sobre a quantidade de pontos e terminais de ônibus perto das instituições de ensino e avaliar se os números e médias de pontos de acesso à transporte público são diferentes para cada região da cidade e índices socioeconômicos dos bairros e escolas.

O segundo objetivo era explorar os dados com objetivo de realizar a representação dos dados de maneira integrada. Ele foi completado construindo uma ontologia para deixar os termos e relações em um mesmo vocabulário. A ontologia construída foi intitulada de *Mobility Education Ontology* (MOBED).

Com a ontologia desenvolvida foi possível completar o terceiro objetivo, conectar os dados abertos referentes à mobilidade e educação da cidade de Curitiba. Os dados presentes nos nas bases de dados utilizados foram mapeados para classes, propriedades de objeto e de dados disponibilizados na ontologia. O mapeamento foi feito através do *framework Ontop*.

O terceiro objetivo se refere a compreender como podem *Linked Open Data* pode ser implementado no contexto da cidade de Curitiba para os domínios de Educação e Mobilidade. Para isso foram utilizadas ferramentas para o tratamento dos dados abertos, ferramentas para a construção da ontologia e para a realização de consultas.

O último objetivo específico deste trabalho é Compreender como *Linked Open Data* podem produzir novas informações relevantes para a cidade de Curitiba. Para isso, as triplas foram materializadas em um arquivo, que foi inserido em uma *Triple Store* chamada *GraphDB*. Ele que possibilitou a criação de *queries* em SPARQL, para a análise dos dados.

Os resultados gerados não são capazes de gerar hipóteses concretas para dizer

que a mobilidade urbana, sem uma análise mais completa envolvendo outros fatores, tem alguma influência direta na educação e na qualidade do ensino das instituições educacionais de Curitiba. Entretanto, organizam os dados para que isso seja possível com a aplicação de alguma técnica estatística ou de análise de dados. Os resultados obtidos dessas análises podem ajudar gestores, pesquisadores e cidadãos a desenvolver novas soluções para o município, ou investigar situações específicas a que tenham interesse.

Além disso, este trabalho demonstra a possibilidade e metodologia para se realizar integração de dados abertos para diferentes domínios de cidades inteligentes e para casos de diferentes cidades e regiões.

Há limitações ao se desenvolver essas relações diretas entre domínios em cidades, como a proposta por este trabalho, principalmente quando estão sendo avaliados apenas mobilidade urbana através de transporte público municipal e índices de desempenho acadêmicos. São elas:

1. O foco deste trabalho de conclusão de curso se encontra na análise sobre a acessibilidade à transporte público. E não há como garantir que todos os dados analisados irão ser referentes a alunos que utilizam o transporte público da cidade.
2. Por mais que a hipótese desenvolvida por este trabalho seja válida, não há garantia que existe uma relação direta entre mobilidade urbana e a qualidade da educação do município.
3. Há outros fatores que poderiam ser inseridos nessa avaliação para haver um visão mais complexa e geral do município: Condição do transporte, condição das vias, existência de ciclovias e uso de aplicativos de transporte privado, por exemplo.

8.1 Trabalhos Futuros

Como trabalhos futuros pretende-se estender a ontologia para um modelo ontológico que seja capaz de interpretar e conectar mais domínios de cidades inteligentes, e conectar com outras ontologias que sejam capazes de tratar outros domínios de cidades inteligentes.

Além disso, pretende-se explorar meios para simplificar o nível técnico exigido para realizar consultas de informações sobre os dados conectados, para que cidadãos também possam desenvolver soluções a partir dos dados dispostos por esta ontologia. E realizar análises estatísticas mais complexas sobre as novas informações obtidas da relação entre os domínios de educação e mobilidade.

8.2 Desdobramentos do Trabalho

Ao longo do desenvolvimento deste trabalho, ele foi compartilhado com a comunidade científica com a intenção de obter críticas construtivas sobre seu desenvolvimento. E

assim, alguns desdobramentos surgiram em congressos e eventos acadêmicos.

Como frutos deste trabalho, foi possível realizar a publicação de um artigo, sobre a revisão sistemática executada, pela *Americas Conference on Information Systems* (AMCIS) com o título de 'Use of Smart and Open Data in Smart Cities'¹ em julho de 2019.

Além disso, a pesquisa teve a honra de receber os títulos de 'Voto Popular' e 'Mérito Acadêmico' na categoria de "Your Research in a Video" da *4th Conference on Information Systems in Latin America* (ISLA 2020) em setembro de 2020 ².

¹Disponível em: https://aisel.aisnet.org/amcis2019/digital_government/digital_government/7/

²Disponível em: https://www.youtube.com/watch?v=pUwkvXXrbNYfeature=youtu.beab_channel=LacaisTube%2FISLA

Referências

- ALAVI, A. H. et al. Internet of things-enabled smart cities: State-of-the-art and future trends. **Measurement**, Elsevier, v. 129, p. 589–606, 2018. Citado na página 25.
- ALAWADHI, S. et al. Building understanding of smart city initiatives. In: SPRINGER. **International conference on electronic government**. [S.l.], 2012. p. 40–53. Citado na página 33.
- ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. **Journal of urban technology**, Taylor & Francis, v. 22, n. 1, p. 3–21, 2015. Citado na página 15.
- ALGEMILI, U. A. Outstanding challenges in recent open government data initiatives. **International Journal of e-Education, e-Business, e-Management and e-Learning**, IACSIT Press, v. 6, n. 2, p. 91, 2016. Citado 2 vezes nas páginas 15 e 26.
- ALIC, A. S. et al. Bigsea: A big data analytics platform for public transportation information. **Future Generation Computer Systems**, Elsevier, v. 96, p. 243–269, 2019. Citado na página 56.
- ARZBERGER, P. et al. **An international framework to promote access to data**. [S.l.]: American Association for the Advancement of Science, 2004. Citado na página 40.
- BALASUBRAMANI, B. S. et al. Ontology-based urban data exploration. In: ACM. **Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics**. [S.l.], 2016. p. 10. Citado na página 24.
- BALESTRINI, M. et al. A city in common: a framework to orchestrate large-scale citizen engagement around urban issues. In: ACM. **Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems**. [S.l.], 2017. p. 2282–2294. Citado na página 26.
- BELIZARIO, M. G.; BERARDI, R. C. Use of smart data and open data in smart cities. **Americas Conference on Information Systems**, Cancún, v. 1, 2019. No prelo. Citado 2 vezes nas páginas 19 e 23.
- BERGMAN, K. **Understanding the relationship between property development and Bus Rapid Transit: A spatiotemporal analysis of transit oriented development in Curitiba, Brazil**. 2017. Citado na página 53.
- BERNERS-LEE, T. **Linked data-design issues**. 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Citado 2 vezes nas páginas 38 e 40.
- BESEISO, M.; AL-ALWANI, A.; ALTAMEEM, A. An interoperable data framework to manipulate the smart city data using semantic technologies. **International Journal of Advanced computer science and applications**, 2017. Citado na página 27.
- BISCHOF, S. et al. Semantic modelling of smart city data. 2014. Citado na página 16.

- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: The story so far. In: **Semantic services, interoperability and web applications: emerging concepts**. [S.l.]: IGI Global, 2011. p. 205–227. Citado 2 vezes nas páginas 37 e 38.
- BOLÍVAR, M. P. R. **Smart Technologies for Smart Governments**. [S.l.]: Springer, 2018. Citado na página 15.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. **Computer networks and ISDN systems**, Elsevier, v. 30, n. 1-7, p. 107–117, 1998. Citado na página 37.
- CHOURABI, H. et al. Understanding smart cities: An integrative framework. In: **IEEE. 2012 45th Hawaii international conference on system sciences**. [S.l.], 2012. p. 2289–2297. Citado na página 16.
- CONSOLI, S. et al. A smart city data model based on semantics best practice and principles. In: **ACM. Proceedings of the 24th International Conference on World Wide Web**. [S.l.], 2015. p. 1395–1400. Citado na página 27.
- CONSOLI, S. et al. An urban fault reporting and management platform for smart cities. In: **ACM. Proceedings of the 24th International Conference on World Wide Web**. [S.l.], 2015. p. 535–540. Citado na página 28.
- DAHL, R. A. A democratic paradox? 1. **Scandinavian Political Studies**, 2000. Citado na página 41.
- DAMBRUCH, J. Computing feedback for citizens' proposals in participative urban planning. **ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 4, 2018. Citado 2 vezes nas páginas 26 e 27.
- DAVIES, J.; STUDER, R.; WARREN, P. **Semantic Web technologies: trends and research in ontology-based systems**. [S.l.]: John Wiley & Sons, 2006. Citado na página 38.
- DAVIES, T. Open data, democracy and public sector reform. **A look at open government data use from data. gov. uk**, 2010. Citado na página 41.
- DEGBELO, A. Linked data and visualization: two sides of the transparency coin. In: **ACM. Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics**. [S.l.], 2017. p. 13. Citado na página 25.
- DUARTE, F. et al. What to expect from the future leaders of bogotá and curitiba in terms of public transport: Opinions and practices among university students. **Transportation Research Part F: Traffic Psychology and Behaviour**, Elsevier, v. 38, p. 7–21, 2016. Citado na página 53.
- DUBOIS, D. L. et al. Prospective investigation of the effects of socioeconomic disadvantage, life stress, and social support on early adolescent adjustment. **Journal of abnormal psychology**, American Psychological Association, v. 103, n. 3, p. 511, 1994. Citado na página 36.
- FORTINO, G. et al. Towards multi-layer interoperability of heterogeneous iot platforms: The inter-iot approach. In: **Integration, interconnection, and interoperability of IoT systems**. [S.l.]: Springer, 2018. p. 199–232. Citado na página 28.

GASTAUD, E. From internet of things to smart data for smart urban monitoring. **International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 42, 2017. Citado na página 27.

GEBETSROITHER-GERINGER, E.; STOLLNBERGER, R.; PETERS-ANDERS, J. Interactive spatialweb-applications as newmeans of support for urban decision-making processes. **ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 4, 2018. Citado na página 25.

GIFFINGER, R.; PICHLER-MILANOVIĆ, N. **Smart cities: Ranking of European medium-sized cities**. [S.l.]: Centre of Regional Science, Vienna University of Technology, 2007. Citado 3 vezes nas páginas 22, 34 e 35.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, Elsevier, v. 5, n. 2, p. 199–220, 1993. Citado na página 39.

GUARINO, N. Understanding, building and using ontologies. **International Journal of Human-Computer Studies**, Elsevier, v. 46, n. 2-3, p. 293–310, 1997. Citado na página 39.

GUARINO, N.; OBERLE, D.; STAAB, S. What is an ontology? In: **Handbook on ontologies**. [S.l.]: Springer, 2009. p. 1–17. Citado 2 vezes nas páginas 16 e 39.

HALL, R. E. et al. **The vision of a smart city**. [S.l.], 2000. Citado na página 34.

HARRISON, C.; DONNELLY, I. A. A theory of smart cities. In: **Proceedings of the 55th Annual Meeting of the ISSS-2011, Hull, UK**. [S.l.: s.n.], 2011. v. 55, n. 1. Citado na página 33.

HARTOG, M.; MULDER, A. A linked open data and semantic citizen vocabulary for public administration: a first exploration and implementation. In: ESOCIETY INSTITUTE OF THE HAGUE UNIVERSITY OF APPLIED SCIENCES. **International Conference Internet Technologies & Society**. [S.l.], 2017. p. 115–125. Citado 2 vezes nas páginas 26 e 27.

HEINLEIN, L. M.; SHINN, M. School mobility and student achievement in an urban setting. **Psychology in the Schools**, Wiley Online Library, v. 37, n. 4, p. 349–357, 2000. Citado 2 vezes nas páginas 17 e 36.

HONARVAR, A. R.; SAMI, A. **A multi-source big data analytic system in smart city for urban planning and decision making**. [S.l.]: GRIN Verlag, 2016. Citado na página 27.

INGERSOLL, G. M.; SCAMMAN, J. P.; ECKERLING, W. D. Geographic mobility and student achievement in an urban setting. **Educational evaluation and policy analysis**, Sage Publications Sage CA: Los Angeles, CA, v. 11, n. 2, p. 143–149, 1989. Citado na página 36.

ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados: Em busca da Web do Conhecimento**. [S.l.]: Novatec Editora, 2015. Citado 5 vezes nas páginas 37, 38, 39, 40 e 41.

- IWANIAK, A. et al. Semantic metadata for heterogeneous spatial planning documents. **ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 3, n. 4, 2016. Citado na página 27.
- JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. Benefits, adoption barriers and myths of open data and open government. **Information systems management**, Taylor & Francis, v. 29, n. 4, p. 258–268, 2012. Citado 3 vezes nas páginas 15, 40 e 41.
- JANSSEN, M.; MATHEUS, R.; ZUIDERWIJK, A. Big and open linked data (bold) to create smart cities and citizens: Insights from smart energy and mobility cases. In: SPRINGER. **International Conference on Electronic Government**. [S.l.], 2015. p. 79–90. Citado 2 vezes nas páginas 30 e 31.
- JOHANSSON, D.; LASSINANTTI, J.; WIBERG, M. Mobile e-services and open data in e-government processes: transforming citizen involvement. In: ACM. **Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services**. [S.l.], 2015. p. 8. Citado na página 25.
- KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004. Citado na página 19.
- KLYNE, G.; CARROLL, J. J. Resource description framework (rdf): Concepts and abstract syntax. 2006. Citado 2 vezes nas páginas 37 e 40.
- KOZIEVITCH, N. P. et al. Exploratory analysis of public transportation data in curitiba. **43o Seminário Integrado de Software e Hardware, Porto Alegre**, p. 1656–1666, 2016. Citado na página 42.
- LEMOS, A. Cidades inteligentes. **GV-executivo**, v. 12, n. 2, p. 46–49, 2013. Citado na página 15.
- LIMA, C. d. A.; FORTUNATO, R. A. Mobilidade e qualidade espacial urbana no entorno de terminais do sistema brt de curitiba: desenho urbano e condições socioambientais. **urbe. Revista Brasileira de Gestão Urbana**, SciELO Brasil, v. 9, p. 329–345, 2017. Citado na página 53.
- LIU, X. et al. Scicloud: A scientific cloud and management platform for smart city data. In: IEEE. **2017 28th International Workshop on Database and Expert Systems Applications (DEXA)**. [S.l.], 2017. p. 27–31. Citado na página 27.
- LONG, L. International perspectives on the residential mobility of america's children. **Journal of Marriage and the Family**, JSTOR, p. 861–869, 1992. Citado na página 17.
- MAEDCHE, A.; STAAB, S. Ontology learning for the semantic web. **IEEE Intelligent systems**, IEEE, v. 16, n. 2, p. 72–79, 2001. Citado na página 39.
- MARCEAU, J. **Introduction: Innovation in the city and innovative cities**. [S.l.]: Taylor & Francis, 2008. Citado na página 15.
- MCMILLAN, D. et al. Data and the city. In: ACM. **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. [S.l.], 2016. p. 2933–2944. Citado 2 vezes nas páginas 24 e 25.

- MEHTA, M. Y.; BUCH, S. Big data mining and semantic technologies: challenges and opportunities. **International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)**, v. 3, n. 7, p. 4907–4913, 2015. Citado na página 24.
- MELLOULI, S.; LUNA-REYES, L. F.; ZHANG, J. Smart government, citizen participation and open data. **Information Polity**, IOS Press, v. 19, n. 1, 2, p. 1–4, 2014. Citado na página 15.
- MOUSTAKA, V.; VAKALI, A.; ANTHOPOULOS, L. G. A systematic review for smart city data analytics. **ACM Computing Surveys (CSUR)**, ACM, v. 51, n. 5, p. 103, 2018. Citado na página 24.
- NAM, T.; PARDO, T. A. Conceptualizing smart city with dimensions of technology, people, and institutions. In: ACM. **Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times**. [S.l.], 2011. p. 282–291. Citado 2 vezes nas páginas 35 e 41.
- NAPHADE, M. et al. Smarter cities and their innovation challenges. **Computer**, IEEE, v. 44, n. 6, p. 32–39, 2011. Citado na página 15.
- NEIROTTI, P. et al. Current trends in smart city initiatives: Some stylised facts. **Cities**, Elsevier, v. 38, p. 25–36, 2014. Citado 2 vezes nas páginas 33 e 34.
- NESI, P. et al. An integrated smart city platform. In: SPRINGER. **Semanitic Keyword-based Search on Structured Data Sources**. [S.l.], 2017. p. 171–176. Citado na página 30.
- NOY, N. F.; MCGUINNESS, D. L. et al. **Ontology development 101: A guide to creating your first ontology**. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05, 2001. Citado na página 44.
- NOY, N. F. et al. Creating semantic web contents with protege-2000. **IEEE intelligent systems**, IEEE, v. 16, n. 2, p. 60–71, 2001. Citado 2 vezes nas páginas 53 e 57.
- PIOVANO, L. et al. What (smart) data visualizations can offer to smart city science. **Communications & Strategies**, n. 96, p. 89, 2014. Citado na página 25.
- RAYCHAUDHURI, A. et al. Factors affecting students' academic performance: A case study in agartala municipal council area. **Bangladesh. e-Journal of Sociology**, v. 7, n. 2, p. 34–41, 2010. Citado na página 36.
- RIOS, P. **Creating "The Smart City"**. Tese (Doutorado), 2012. Citado 2 vezes nas páginas e 34.
- ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. **Expert systems with applications**, Elsevier, v. 33, n. 1, p. 135–146, 2007. Citado na página 31.
- SANSEVERINO, E. R.; SANSEVERINO, R. R.; ANELLO, E. A cross-reading approach to smart city: A european perspective of chinese smart cities. **Smart Cities**, Multidisciplinary Digital Publishing Institute, v. 1, n. 1, p. 26–52, 2018. Citado na página 24.

- SCHAFFERS, H. et al. Smart cities and the future internet: Towards cooperation frameworks for open innovation. In: SPRINGER. **The future internet assembly**. [S.l.], 2011. p. 431–446. Citado na página 16.
- SEABORNE, A.; PRUD’HOMMEAUX, E. **SPARQL Query Language for RDF**. [S.l.], 2008. [Http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/](http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/). Citado na página 38.
- STANLEY, D. et al. The usefulness of social exclusion as a theoretical concept to inform social policy in transport. 2007. Citado na página 35.
- THOMPSON, E. M. Smart city: Adding to the complexity of cities. University of Oulu, 2016. Citado na página 24.
- WASHBURN, D. et al. Helping cities understand “smart city” initiatives. **Growth**, v. 17, n. 2, p. 1–17, 2009. Citado na página 34.
- WILLIAMSON, B. Smarter learning software: Education and the big data imaginary. In: **Big Data—Social Data**. [S.l.: s.n.], 2015. Citado na página 25.
- ZANELLA, A. et al. Internet of things for smart cities. **IEEE Internet of Things journal**, IEEE, v. 1, n. 1, p. 22–32, 2014. Citado na página 15.

Apêndices

APÊNDICE A – Formulário: Mobilidade e Educação em Curitiba

Este formulário faz parte do trabalho de conclusão de curso intitulado "Modelo Ontológico para a integração semântica de dados em Cidades Inteligentes: Mobilidade e educação em Curitiba" da Universidade Tecnológica Federal do Paraná (UTFPR). O trabalho tem como objetivo realizar a integração semântica de dados de Mobilidade e Educação da cidade de Curitiba para compreender os principais problemas, necessidades e possíveis soluções que possam ajudar gestores, cientistas e cidadãos

Este formulário tem como objetivo identificar quais são as percepções sobre a relação entre Mobilidade Urbana e Educação sobre a perspectiva de especialistas destes domínios. Os dados coletados por este formulário serão usados para entender como cada um desses domínios implica no outro e para entender como a integração de seus dados pode trazer benefícios para ambos domínios.

29/04/2020

Mobilidade e Educação em Curitiba

Mobilidade e Educação em Curitiba

Este formulário faz parte do trabalho de conclusão de curso intitulado "Modelo Ontológico para a integração semântica de dados em Cidades Inteligentes: Mobilidade e educação em Curitiba" da Universidade Tecnológica Federal do Paraná (UTFPR). O trabalho tem como objetivo realizar a integração semântica de dados de Mobilidade e Educação da cidade de Curitiba para compreender os principais problemas, necessidades e possíveis soluções que possam ajudar gestores, cientistas e cidadãos

Este formulário tem como objetivo identificar quais são as percepções sobre a relação entre Mobilidade Urbana e Educação sobre a perspectiva de especialistas destes domínios. Os dados coletados por este formulário serão usados para entender como cada um desses domínios implica no outro e para entender como a integração de seus dados pode trazer benefícios para ambos domínios.

***Obrigatório**

1. Como você se identifica? Responda de acordo com o domínio em que você trabalha. *

Marcar apenas uma oval.

- Especialista no domínio da educação *Pular para a pergunta 2*
- Especialista no domínio da mobilidade *Pular para a pergunta 18*
- Estudante *Pular para a pergunta 38*

Seção do especialista da educação

Responda de acordo com sua formação e/ou vivência profissional.

2. Você mora em Curitiba? *

Marcar apenas uma oval.

- Sim
- Não

29/04/2020

Mobilidade e Educação em Curitiba

3. Qual a sua relação com o domínio da educação?

Marcar apenas uma oval.

- Sou professor(a) ensino fundamental
- Sou professor(a) ensino médio
- Sou professor pesquisador da área
- Sou gestor(a) de escola/universidade..
- Outro: _____

4. Como você avalia o transporte público na cidade de Curitiba? *

Marcar apenas uma oval.

- | | | | | | | |
|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------|
| | 1 | 2 | 3 | 4 | 5 | |
| Péssimo | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Ótimo |

5. Você acredita que existe alguma relação entre a mobilidade urbana e a educação na cidade de Curitiba? *

Marcar apenas uma oval.

- Sim
- Não
- Talvez
- Não sei

6. Justifique a resposta anterior. *

29/04/2020

Mobilidade e Educação em Curitiba

7. A longa distância das escolas, universidades e cursos pode gerar uma instabilidade no desempenho acadêmico dos alunos? *

Marcar apenas uma oval.

- Sim
 Não
 Talvez
 Não sei

8. Justifique a resposta anterior. *

9. O sistema de planejamento urbano da cidade e as linhas de transporte público atualmente têm facilitado o acesso aos centros de educação (escolas/universidades/cursos) em Curitiba? *

Marcar apenas uma oval.

	1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

10. Justifique a resposta anterior. *

29/04/2020

Mobilidade e Educação em Curitiba

11. Dentre as vulnerabilidades a seguir envolvendo mobilidade, quais podem estar relacionadas à vulnerabilidade acadêmica? *

Marque todas que se aplicam.

- Distância da escola/universidade/curso
- Preço para deslocamento (tarifas de ônibus, táxis, aplicativos)
- Ausência de linhas de transporte público coerentes com o endereço da escola/universidade/curso
- Tempo de deslocamento para chegar à escola/universidade/cursos
- Horários incompatíveis com os horários do início e fim das aulas
- Super lotação
- Transporte sem acessibilidade para pessoas com deficiências
- Nenhuma das anteriores

Outro: _____

12. O quanto você acredita que a melhoria da mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), estaria relacionada a uma melhoria no desempenho acadêmico das escolas e universidades? *

Marcar apenas uma oval.

	1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

13. O quanto você acredita que uma melhoria na mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), estaria relacionada a uma possível diminuição do nível de evasão em todos os níveis da educação? *

Marcar apenas uma oval.

	1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

29/04/2020

Mobilidade e Educação em Curitiba

14. Uma visão integrada desses dois domínios, educação e mobilidade pode trazer benefícios para o desenvolvimento da cidade? *

Marcar apenas uma oval.

	1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

15. Justifique a resposta anterior. *

16. Deixe alguma consideração sobre a relação entre Mobilidade e Educação de acordo com sua vivência e experiência.

17. Quais informações relacionadas à mobilidade urbana você gostaria de ter dos gestores de mobilidade urbana que auxiliariam na gestão da educação em uma cidade?

29/04/2020

Mobilidade e Educação em Curitiba

Seção do especialista de
Mobilidade

Responda de acordo com sua formação e/ou vivência
profissional.

18. Você mora em Curitiba?

Marcar apenas uma oval.

Sim

Não

19. Qual a sua atividade com relação à Mobilidade?

Marcar apenas uma oval.

Trabalho na área de Arquitetura

Trabalho com Planejamento urbano

Sou acadêmico desta área

Outro: _____

20. Você acredita que existe alguma relação entre a mobilidade urbana e a educação na cidade de Curitiba? *

Marcar apenas uma oval.

Sim

Não

Talvez

Não sei

29/04/2020

Mobilidade e Educação em Curitiba

21. Justifique a resposta anterior. *

22. Como você avalia o transporte público na cidade de Curitiba? *

Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ótimo

23. O sistema de planejamento urbano da cidade e as linhas de transporte público atualmente têm facilitado o acesso aos centros de educação (escolas/universidades/cursos) em Curitiba? *

Marcar apenas uma oval.

	1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

24. Justifique a resposta anterior. *

29/04/2020

Mobilidade e Educação em Curitiba

25. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino fundamental?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Importante

26. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino Médio?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Importante

27. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino Superior?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Importante

28. Dentre os aspectos a seguir sobre Educação, quais você acredita serem impactados por conta da qualidade da Mobilidade na cidade de Curitiba? *

Marque todas que se aplicam.

- Evasão em todos os níveis (infantil, médio e superior)
- Número de presença dos alunos em aulas
- Nível de atenção dos alunos em aulas
- Número de atrasos dos alunos para chegar as aulas no horário
- Desempenho acadêmico (notas, trabalhos e avaliações)
- Nenhuma das anteriores

Outro: _____

29/04/2020

Mobilidade e Educação em Curitiba

29. Melhorar a mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), poderia ter um impacto positivo na Educação? *

Marcar apenas uma oval.

1	2	3	4	5		
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

30. Melhorar a mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), poderia melhorar a frequência de presenças e diminuir o nível de evasão em todos os níveis da educação? *

Marcar apenas uma oval.

1	2	3	4	5		
Não, defintivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, defintivamente

31. Uma visão integrada desses dois domínios, educação e mobilidade pode trazer benefícios para o desenvolvimento da cidade? *

Marcar apenas uma oval.

1	2	3	4	5		
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

32. Justifique a resposta anterior. *

29/04/2020

Mobilidade e Educação em Curitiba

33. Quais são os principais terminais de ônibus utilizados por estudantes na cidade de Curitiba?

34. Quais são as principais linhas de ônibus utilizadas por estudantes na cidade de Curitiba?

35. Deixe alguma consideração sobre a relação entre Mobilidade e Educação de acordo com sua vivência e experiência.

36. Qual pergunta você gostaria de fazer para um especialista de educação no que diz respeito à mobilidade urbana na cidade de Curitiba?

29/04/2020

Mobilidade e Educação em Curitiba

37. Quais informações relacionadas à educação você gostaria de ter de forma que auxiliaria na atividade de gestão urbana no quesito mobilidade?

Seção de Estudantes

Responda de acordo com a sua vivência.

38. Em que nível de educação você se encontra no momento

Marcar apenas uma oval.

- Ensino Fundamental
- Ensino médio
- Ensino Superior
- Outro: _____

39. Como você se desloca para chegar à centros de educação (escolas/universidades/cursos) em Curitiba?

Marque todas que se aplicam.

- A pé
- Carro
- Bicicleta
- Aplicativos de transporte privado (UBER, 99POP, etc..)
- Transporte público
- Outro: _____

29/04/2020

Mobilidade e Educação em Curitiba

40. Você acredita que existe alguma relação entre a mobilidade urbana e a educação na cidade de Curitiba?

Marcar apenas uma oval.

- Sim
 Não
 Talvez
 Não Sei

41. Como você avalia o transporte público na cidade de Curitiba?

Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ótimo

42. Justifique a resposta anterior.

43. O sistema de planejamento urbano da cidade e as linhas de transporte público atualmente têm facilitado o acesso aos centros de educação (escolas/universidades/cursos) em Curitiba?

Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ótimo

29/04/2020

Mobilidade e Educação em Curitiba

44. Justifique a resposta anterior.

45. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino fundamental?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Relevante

46. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino Médio?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Relevante

47. Quão importante considera o transporte público para a Mobilidade de estudantes do ensino Superior?

Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito Relevante

29/04/2020

Mobilidade e Educação em Curitiba

48. Dentre as vulnerabilidades a seguir envolvendo mobilidade, quais já afetaram seu desempenho acadêmico (notas, presenças, atrasos, avaliações e nível de atenção em sala de aula)?

Marque todas que se aplicam.

- Distância da escola/universidade/curso
- Preço para deslocamento (tarifas de ônibus, táxis, aplicativos)
- Ausência de linhas de transporte público coerentes com o endereço da escola/universidade/curso
- Tempo de deslocamento para chegar à escola/universidade/cursos
- Horários incompatíveis com os horários do início e fim das aulas
- Super lotação
- Transporte sem acessibilidade para pessoas com deficiências
- Nenhuma das anteriores

Outro: _____

49. Melhorar a mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), poderia ter um impacto positivo na Educação?

Marcar apenas uma oval.

1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

50. Melhorar a mobilidade da cidade em algum aspecto, (preço, tempo, distância, acessibilidade e opções de transporte), poderia melhorar a frequência de presenças e diminuir o nível de evasão em todos os níveis da educação?

Marcar apenas uma oval.

1	2	3	4	5	
Não, definitivamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sim, definitivamente

29/04/2020

Mobilidade e Educação em Curitiba

51. Deixe alguma consideração sobre a relação entre Mobilidade e Educação de acordo com sua vivência e experiência.

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE B – Código: Tratamento de Dados Prova Brasil 2017

```

1 import pandas as pd
2
3 def treatValueINSE(stringValue):
4     if stringValue == 'Muito-baixo':
5         return 1
6     elif stringValue == 'Baixo':
7         return 2
8     elif stringValue == 'M dio-baixo':
9         return 3
10    elif stringValue == 'M dio':
11        return 4
12    elif stringValue == 'M dio-alto':
13        return 5
14    elif stringValue == 'Alto':
15        return 6
16    elif stringValue == 'Muito-alto':
17        return 7
18
19 def main():
20     global csvFile
21
22     #PB_5_MAT
23     csvFile = pd.read_csv(filepath_or_buffer="prova_brasil_5_mat_2017.csv")
24
25     #Limpar dados inconsistentes
26     csvFile = csvFile[csvFile['NSE* (Classe)'] != 'Sem dados*']
27     csvFile = csvFile[csvFile['Percentual Proficiente (%)'] != 'Sem dados*']
28
29     #Tratar valor do INSE
30     csvFile['NSE* (Classe)'] = csvFile['NSE* (Classe)'].apply(treatValueINSE)
31
32     #EXPORT
33     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\prova_brasil_5_mat_2017.csv', index
34                   =None, header=True)
35
36     #-----
37
38     #PB_5_PT
39     csvFile = pd.read_csv(filepath_or_buffer="prova_brasil_5_port_2017.csv")
40
41     #Limpar dados inconsistentes
42     csvFile = csvFile[csvFile['NSE* (Classe)'] != 'Sem dados*']
43     csvFile = csvFile[csvFile['Percentual Proficiente (%)'] != 'Sem dados*']
44
45     #Tratar valor do INSE
46     csvFile['NSE* (Classe)'] = csvFile['NSE* (Classe)'].apply(treatValueINSE)
47
48     #EXPORT
49     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\prova_brasil_5_port_2017.csv',
50                   index=None, header=True)
51
52     #-----

```

```
52  #PB_9_MAT
53  csvFile = pd.read_csv(filepath_or_buffer="prova_brasil_9_mat_2017.csv")
54
55  #Limpar dados inconsistentes
56  csvFile = csvFile[csvFile['NSE* (Classe)'] != 'Sem dados*']
57  csvFile = csvFile[csvFile['Percentual Proficiente (%)'] != 'Sem dados*']
58
59  #Tratar valor do INSE
60  csvFile['NSE* (Classe)'] = csvFile['NSE* (Classe)'].apply(treatValueINSE)
61
62  #EXPORT
63  csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\prova_brasil_9_mat_2017.csv', index
64              =None, header=True)
65
66
67  #PB_9_PT
68  csvFile = pd.read_csv(filepath_or_buffer="prova_brasil_9_port_2017.csv")
69
70  #Limpar dados inconsistentes
71  csvFile = csvFile[csvFile['NSE* (Classe)'] != 'Sem dados*']
72  csvFile = csvFile[csvFile['Percentual Proficiente (%)'] != 'Sem dados*']
73
74  #Tratar valor do INSE
75  csvFile['NSE* (Classe)'] = csvFile['NSE* (Classe)'].apply(treatValueINSE)
76
77  #EXPORT
78  csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\prova_brasil_9_port_2017.csv',
79              index=None, header=True)
80  if __name__ == '__main__':
81      main()
```

APÊNDICE C – Código: Tratamento de Dados ENEM 2015

```
1 import pandas as pd
2
3 def treatValueINSE(stringValue):
4     return stringValue[6]
5
6 def main():
7     global csvFile
8
9     #Abrir arquivo
10    csvFile = pd.read_csv(filepath_or_buffer="enem_2005_2015.csv", sep=';', engine='
        python')
11
12    #Tirar linhas com INSE null
13    csvFile = csvFile[pd.notnull(csvFile['INSE'])]
14
15    #Filtrar por ano
16    csvFile = csvFile[csvFile['NU_ANO'] == 2015]
17
18    #Tratar valor do INSE
19    csvFile['INSE'] = csvFile['INSE'].apply(treatValueINSE)
20
21    #EXPORT
22    csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\enem_2005_2015.csv', index=None,
        header=True)
23
24 if __name__ == '__main__':
25    main()
```

APÊNDICE D – Código: Tratamentos de Dados ENADE 2017

```

1 import pandas as pd
2 import numpy as np
3 import math
4
5 def exportCSV(csvFile):
6     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\enade_2017.csv', index=None, header
7         =True)
8
9 def main():
10     global csvEnade
11     global csvPonto
12
13     # Abrir arquivo
14     csvEnade = pd.read_csv(filepath_or_buffer="enade_2017.CSV", sep=';', engine='python')
15
16     # Valores iniciais da tabela
17     ult_curso = 849
18     ult_uni = 10
19
20     media_nota = float(0)
21     qt_linhas = float(1)
22
23     nu_ano = []
24     co_ies = []
25     co_curso = []
26     nt_geral = []
27
28     for index, row in csvEnade.iterrows():
29         ies = row['CO_IES']
30         curso = row['CO_CURSO']
31         nota = row['NT_GER']
32
33         if ies != ult_uni or curso != ult_curso:
34             nu_ano.append(2017)
35             co_ies.append(ult_uni)
36             co_curso.append(ult_curso)
37             nt_geral.append(float(media_nota)/float(qt_linhas))
38
39             ult_curso = curso
40             ult_uni = ies
41
42             nota = str(nota).replace(',','.')
43             if nota != 'nan':
44                 if not math.isnan(float(nota)):
45                     media_nota = float(nota)
46                     qt_linhas = 1
47                 else:
48                     media_nota = float(0)
49                     qt_linhas = 1
50             else:
51                 nota = str(nota).replace(',','.')
52                 if nota != 'nan':
53                     if not math.isnan(float(nota)):
54                         media_nota += float(nota)
55                         qt_linhas += 1

```

```
55
56     # dictionary of lists
57     dict = {'ano': nu_ano, 'cod_ies': co_ies, 'co_curso': co_curso, 'nt_geral': nt_geral}
58     df = pd.DataFrame(dict)
59
60     # saving the dataframe
61     # print(df);
62     exportCSV(df)
63
64 if __name__ == '__main__':
65     main()
```

APÊNDICE E – Código: Tratamento de Relação Escolas-Terminais

```

1 import pandas as pd
2 from math import sin, cos, sqrt, atan2, radians
3
4 def exportCSV(csvFile):
5     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\escola_terminal.csv', index=None,
6                   header=True)
7
8 def calculateDistance(lat_a, lat_b, lon_a, lon_b):
9     lat_a = str(lat_a).replace(',', '.', '')
10    lat_b = str(lat_b).replace(',', '.', '')
11    lon_a = str(lon_a).replace(',', '.', '')
12    lon_b = str(lon_b).replace(',', '.', '')
13
14    R = 6373.0
15    lat1 = radians(float(lat_a))
16    lon1 = radians(float(lon_a))
17    lat2 = radians(float(lat_b))
18    lon2 = radians(float(lon_b))
19
20    dlon = lon2 - lon1
21    dlat = lat2 - lat1
22
23    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
24    c = 2 * atan2(sqrt(a), sqrt(1 - a))
25
26    distance = R * c
27
28    if distance < 0.25:
29        return 'PERTO'
30    elif distance < 0.50:
31        return 'MEDIO'
32    else:
33        return 'LONGE'
34
35 def main():
36     global csvEscola
37     global csvPonto
38
39     # Abrir arquivo
40     csvEscola = pd.read_csv(filepath_or_buffer="escolas_curitiba.csv", sep=';', engine='python')
41     csvTerminal = pd.read_csv(filepath_or_buffer="terminal_de_transporte.csv", sep=',', engine='python')
42
43     escola_id = []
44     dist = []
45     terminal_id = []
46
47     # Calcular a distância entre escola e ponto
48     for index, row in csvEscola.iterrows():
49         escola = row['CD_EQUI']
50         lat_a = row['LAT_SIRGAS']
51         long_a = row['LON_SIRGAS']
52
53         for index_2, row_2 in csvTerminal.iterrows():

```

```
53     terminal = row_2['id']
54     lat_b = row_2['lat']
55     long_b = row_2['lng']
56
57     distance = calculateDistance(lat_a, lat_b, long_a, long_b)
58
59     if distance == 'PERTO':
60         #Append data to arrays
61         escola_id.append(escola)
62         dist.append(distance)
63         terminal_id.append(terminal)
64
65     # dictionary of lists
66     dict = {'escola': escola_id, 'distancia': dist, 'terminal': terminal_id}
67     df = pd.DataFrame(dict)
68     # saving the dataframe
69     # print(df);
70     exportCSV(df)
71
72 if __name__ == '__main__':
73     main()
```


APÊNDICE F – Código: Tratamento de Relação Escolas-Pontos de Ônibus

```

1 import pandas as pd
2 from math import sin, cos, sqrt, atan2, radians
3
4 def exportCSV(csvFile):
5     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\escola_ponto.csv', index=None,
6                   header=True)
7
8 def calculateDistance(lat_a, lat_b, lon_a, lon_b):
9     lat_a = str(lat_a).replace(',', '.', '')
10    lat_b = str(lat_b).replace(',', '.', '')
11    lon_a = str(lon_a).replace(',', '.', '')
12    lon_b = str(lon_b).replace(',', '.', '')
13
14    R = 6373.0
15    lat1 = radians(float(lat_a))
16    lon1 = radians(float(lon_a))
17    lat2 = radians(float(lat_b))
18    lon2 = radians(float(lon_b))
19
20    dlon = lon2 - lon1
21    dlat = lat2 - lat1
22
23    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
24    c = 2 * atan2(sqrt(a), sqrt(1 - a))
25
26    distance = R * c
27
28    if distance < 0.25:
29        return 'PERTO'
30    elif distance < 0.50:
31        return 'MEDIO'
32    else:
33        return 'LONGE'
34
35 def main():
36     global csvEscola
37     global csvPonto
38
39     # Abrir arquivo
40     csvEscola = pd.read_csv(filepath_or_buffer="escolas_curitiba.csv", sep=';', engine='
41                             python')
42     csvPonto = pd.read_csv(filepath_or_buffer="pontos_curitiba.csv", sep=',', engine='
43                             python')
44
45     escola_id = []
46     dist = []
47     ponto_id = []
48
49     # Calcular a distância entre escola e ponto
50     for index, row in csvEscola.iterrows():
51         escola = row['CD_EQUI']
52         lat_a = row['LAT_SIRGAS']
53         lon_a = row['LON_SIRGAS']

```

```
51
52     for index_2, row_2 in csvPonto.iterrows():
53         ponto = row_2['NUM']
54         lat_b = row_2['LAT']
55         long_b = row_2['LON']
56
57         distance = calculateDistance(lat_a, lat_b, long_a, long_b)
58
59         if distance == 'PERTO':
60             #Append data to arrays
61             escola_id.append(escola)
62             dist.append(distance)
63             ponto_id.append(ponto)
64
65     # dictionary of lists
66     dict = {'escola': escola_id, 'distancia': dist, 'ponto': ponto_id}
67     df = pd.DataFrame(dict)
68     # saving the dataframe
69     # print(df);
70     exportCSV(df)
71
72 if __name__ == '__main__':
73     main()
```

APÊNDICE G – Código: Tratamento de Relação Universidade-Terminal

```

1 import pandas as pd
2 from math import sin, cos, sqrt, atan2, radians
3
4 def exportCSV(csvFile):
5     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\universidade_terminal.csv', index=
6         None, header=True)
7
8 def calculateDistance(lat_a, lat_b, lon_a, lon_b):
9     lat_a = str(lat_a).replace(',', '.', '')
10    lat_b = str(lat_b).replace(',', '.', '')
11    lon_a = str(lon_a).replace(',', '.', '')
12    lon_b = str(lon_b).replace(',', '.', '')
13
14    R = 6373.0
15    lat1 = radians(float(lat_a))
16    lon1 = radians(float(lon_a))
17    lat2 = radians(float(lat_b))
18    lon2 = radians(float(lon_b))
19
20    dlon = lon2 - lon1
21    dlat = lat2 - lat1
22
23    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
24    c = 2 * atan2(sqrt(a), sqrt(1 - a))
25
26    distance = R * c
27
28    if distance < 0.25:
29        return 'PERTO'
30    elif distance < 0.50:
31        return 'MEDIO'
32    else:
33        return 'LONGE'
34
35 def main():
36     global csvEscola
37     global csvPonto
38
39     # Abrir arquivo
40     csvUniversidade = pd.read_csv(filepath_or_buffer="universidades_curitiba.csv", sep=';',
41         ', ', engine='python')
42     csvTerminal = pd.read_csv(filepath_or_buffer="terminal_de_transporte.csv", sep=',',
43         ', ', engine='python')
44
45     universidade_id = []
46     campus_id = []
47     dist = []
48     terminal_id = []
49
50     # Calcular a distância entre escola e ponto
51     for index, row in csvUniversidade.iterrows():
52         universidade = row['COD_IES']
53         id_uni = row['CD_EQUI']
54         lat_a = row['LAT_SIRGAS']
55         long_a = row['LON_SIRGAS']

```

```
53
54     for index_2, row_2 in csvTerminal.iterrows():
55         terminal = row_2['id']
56         lat_b = row_2['lat']
57         long_b = row_2['lng']
58
59         distance = calculateDistance(lat_a, lat_b, long_a, long_b)
60
61         if distance == 'PERTO':
62             #Append data to arrays
63             universidade_id.append(universidade)
64             campus_id.append(id_uni)
65             dist.append(distance)
66             terminal_id.append(terminal)
67
68     # dictionary of lists
69     dict = {'campus': campus_id, 'universidade': universidade_id, 'distancia': dist, '
70           terminal': terminal_id}
71     df = pd.DataFrame(dict)
72     # saving the dataframe
73     # print(df);
74     exportCSV(df)
75 if __name__ == '__main__':
76     main()
```

APÊNDICE H – Código: Tratamento de Relação Universidade-Ponto de Ônibus

```

1 import pandas as pd
2 from math import sin, cos, sqrt, atan2, radians
3
4 def exportCSV(csvFile):
5     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\universidade_ponto.csv', index=None
6         , header=True)
7
8 def calculateDistance(lat_a, lat_b, lon_a, lon_b):
9     lat_a = str(lat_a).replace(',','.')
10    lat_b = str(lat_b).replace(',','.')
11    lon_a = str(lon_a).replace(',','.')
12    lon_b = str(lon_b).replace(',','.')
13
14    R = 6373.0
15    lat1 = radians(float(lat_a))
16    lon1 = radians(float(lon_a))
17    lat2 = radians(float(lat_b))
18    lon2 = radians(float(lon_b))
19
20    dlon = lon2 - lon1
21    dlat = lat2 - lat1
22
23    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
24    c = 2 * atan2(sqrt(a), sqrt(1 - a))
25
26    distance = R * c
27
28    if distance < 0.25:
29        return 'PERTO'
30    elif distance < 0.50:
31        return 'MEDIO'
32    else:
33        return 'LONGE'
34
35 def main():
36     global csvEscola
37     global csvPonto
38
39     # Abrir arquivo
40     csvUniversidade = pd.read_csv(filepath_or_buffer="universidades_curitiba.csv", sep=';
41         ', engine='python')
42     csvPonto = pd.read_csv(filepath_or_buffer="pontos_curitiba.csv", sep=',', engine='
43         python')
44
45     universidade_id = []
46     campus_id = []
47     dist = []
48     ponto_id = []
49
50     # Calcular a distância entre escola e ponto
51     for index, row in csvUniversidade.iterrows():
52         id_uni = row['CD_EQUI']
53         universidade = row['COD_IES']

```

```
51     lat_a = row[ 'LAT_SIRGAS' ]
52     long_a = row[ 'LON_SIRGAS' ]
53
54     for index_2, row_2 in csvPonto.iterrows():
55         ponto = row_2[ 'NUM' ]
56         lat_b = row_2[ 'LAT' ]
57         long_b = row_2[ 'LON' ]
58
59         distance = calculateDistance(lat_a, lat_b, long_a, long_b)
60
61         if distance == 'PERTO':
62             #Append data to arrays
63             universidade_id.append(universidade)
64             campus_id.append(id_uni)
65             dist.append(distance)
66             ponto_id.append(ponto)
67
68     # dictionary of lists
69     dict = { 'campus': campus_id, 'universidade': universidade_id, 'distancia': dist, '
70             ponto': ponto_id }
71     df = pd.DataFrame(dict)
72     # saving the dataframe
73     # print(df);
74     exportCSV(df)
75 if __name__ == '__main__':
76     main()
```

APÊNDICE I – Código: Tratamento de Pontos de Ônibus de Curitiba

```
1 import pandas as pd
2 import requests
3 import json
4
5 def exportCSV(csvFile):
6     csvFile.to_csv(r'C:\Users\Mateus\Desktop\Tratados\escola_ponto.csv', index=None,
7                   header=True)
8
9 def main():
10     global csvLinhas
11
12     # Abrir arquivo
13     csvLinhas = pd.read_csv(filepath_or_buffer="linha_onibus.csv", sep=',', engine='
14     python')
15     jsonFile = '';
16
17     for index, row in csvLinhas.iterrows():
18         cod = row['cod_linha']
19         response = requests.get("http://transporteservico.urbs.curitiba.pr.gov.br/
20         getPontosLinha.php?linha="+cod+"&c=e6244")
21
22         responseJson = response.json()
23         if (len(responseJson) > 0):
24
25             for x in range(len(responseJson)):
26                 responseJson[x]['COD_LINHA'] = cod
27
28             if (index == 0):
29                 jsonFile = json.dumps(responseJson)
30             else:
31                 res_dict = jsonFile + json.dumps(responseJson)
32                 jsonFile = res_dict
33
34         f = open("pontos_linhas.json", "a")
35         f.write(jsonFile)
36         f.close()
37
38 if __name__ == '__main__':
39     main()
```