

Uma discussão sobre a relação entre Segurança e Desempenho Acadêmico de estudantes do ensino médio apoiada por dados abertos conectados de Curitiba

Pedro Henrique S. Auceli, Mateus G. Belizario, Nádia P. Kozievitch, Rita C. G. Berardi,

Departamento Acadêmico de Informática
Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba, PR – Brazil
{pedroauceli, mateusbelizario}@alunos.utfpr.edu.br, {ritaberardi, nadiap}@utfpr.edu.br

***Abstract.** The objective of this work is to show how to use linked open data to support the discussion about the relationship between education and other social domains. The domain chosen for a first discussion using these data was that of security due to the hypothesis that the social environment in which the school is inserted can influence academic performance and even student dropout. With open data from the city of Curitiba it was possible to notice some differences in the approval and dropout rates in more and less insecure neighborhoods.*

***Resumo.** O objetivo deste trabalho é mostrar como utilizar dados abertos conectados como apoio na discussão sobre a relação entre educação e outros domínios sociais. O domínio escolhido para uma primeira discussão foi o da segurança, pela hipótese de que o ambiente social em que a escola está inserida pode influenciar no desempenho acadêmico e até evasão de estudantes. Com dados abertos da cidade de Curitiba foi possível notar algumas diferenças nos índices de aprovação e evasão em escolas de bairros mais e menos inseguros.*

1. Introdução

Com o movimento mundial de abertura de dados governamentais, diversas áreas da sociedade têm se beneficiado com o acesso aos dados por viabilizar um acompanhamento transparente dos serviços públicos. Apesar de o Brasil estar relativamente bem colocado mundialmente¹ quanto à abertura de dados, o uso desses para melhoria de gestão específica para a área da educação ainda pode ser considerado relativamente novo (Bandeira et al., 2015). O estudo de Santos et al. (2017), mostra que foi a partir de 2013 que a comunidade de Informática na Educação no Brasil começou a se interessar pela área de gestão de educação baseada em dados. Vários estudos fazem uso dos dados abertos educacionais com o enfoque direto na prática pedagógica, seja para desenvolvimento de aplicações para dados conectados (Sarker e Farhana, 2014) (Isotani e Bittencourt, 2015) ou para uso dos dados como recurso educacional em sala de aula (Atenas e Havemann, 2015). No entanto, a Informática na Educação, através de análises de dados abertos, também tem se mostrado como uma importante ferramenta no auxílio à gestão da educação. Exemplo disso são trabalhos que utilizam microdados

¹ <https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf>

do censo escolar, como Ferreira (2015) que busca identificar a influência de características familiares no Ensino Fundamental, como também Carvalho et al. (2017) que analisam perfis de instituições e estudantes no Ensino Básico e Superior, Rigo et al. (2012) que estudam comportamentos relacionados à evasão e Adeodato et al. (2014) que estudam a predição de desempenho de escolas.

Uma importante vantagem que os dados abertos trazem é a possibilidade de gerar visões integradas de domínios sociais que na dinâmica original da cidade são inter-relacionados, porém a coleta e processamento de dados acontece de maneira isolada. Isto acontece devido aos diferentes Sistemas de Informação e Bancos de Dados que são projetados para atender cada setor diferentemente. A integração desses dados não é uma tarefa trivial devido a vários fatores, sendo a diversidade semântica um dos que pode ser destacado. Cada domínio possui seus termos e jargões que refletem nas estruturas dos seus dados. Além disso, o formato de disponibilização dos dados é bastante determinante para que o acesso a esses dados aconteça da maneira mais abrangente possível, tanto por pessoas quanto por máquinas. Se os dados são abertos em formatos processáveis por agentes de softwares, seu acesso para cruzamentos com outros dados, análises e visualizações pode aumentar a capacidade de tomada de decisão em problemas complexos por criadores de políticas públicas, pesquisadores e cidadãos (Penteado et al. 2017).

Uma visão integrada de dados da educação com dados de outros domínios sociais pode auxiliar desde professores, gestores de escolas até administradores públicos a criarem dinâmicas de enfrentamento a fatores que impactam na vida escolar dos estudantes. Neste estudo o domínio escolhido para a integração foi o da segurança pela hipótese de que o ambiente social em que uma escola está inserida pode impactar no desempenho e até evasão de estudantes. Tendo em vista que a escola é, em certo sentido, um recorte das turbulências sociais com estudantes em risco de reprovação, evasão, delinquência e violência no ambiente acadêmico. Para compreender tais fenômenos é importante observar vários fatores, inclusive os que são externos ao ambiente escolar mas que podem impactar nesses indicadores. A convivência de estudantes com o clima de violência pode ser um desses fatores que afeta de alguma maneira a vida na escola (Pino, 2007).

Diante deste cenário, este trabalho apresenta uma integração semântica entre o domínio de educação e o domínio de segurança através do uso de tecnologias de *Linked Data* (dados ligados ou dados conectados). A W3C² traz um conjunto de práticas para publicação e uso de dados na web para fazer com o que os dados sejam encontrados e compreendidos tanto por humanos quanto por computadores. As tecnologias de *Linked Data* e *Web Semântica* (Berners-Lee, 2001) (Bizer e Seaborne, 2004) possibilitam tirar proveito adequado da web padronizando vocabulários e formato de dados lidando com a heterogeneidade dos formatos com que os dados são comumente publicados, integrando dados de fontes diferentes, cada qual com sua própria frequência de amostragem, características, formatos e semântica. Essas tecnologias permitem a criação e publicação de uma visão semântica única, chamada *Linked Data Mashup* (LDM) (Avila et al. 2018). O *mashup* foi criado neste estudo com o objetivo de integrar os dados semanticamente para facilitar a discussão sobre as possíveis relações entre os domínios educação e segurança. Para mostrar como o *mashup* pode ser utilizado para

² <https://www.w3.org/TR/dwbp>

apoiar tal discussão foram definidas três questões de competência: i) *Existe uma relação entre o rendimento dos alunos e a quantidade de alunos por turma, com a quantidade de atendimentos feitos pela Guarda Municipal no bairro em que a escola se encontra?*; ii) *Existe uma diferença entre o rendimento e o tamanho das turmas entre escolas públicas e particulares de um bairro?* e iii) *Existe uma relação entre a taxa de abandono e o volume de atendimentos pela Guarda Municipal nos bairros das escolas?*

Os dados utilizados foram os dados abertos referentes à cidade de Curitiba, capital do estado do Paraná. Acredita-se que apesar dos dados se tratarem de uma única cidade brasileira, o trabalho traz uma contribuição para que outras cidades se utilizem do modelo de integração, seja adotando a metodologia descrita ou até mesmo na utilização do *mashup* criado neste estudo para possíveis comparações com outras cidades.

2. Trabalhos Relacionados

O trabalho de Ávila et al. (2018) utilizou técnicas de *Linked Data* e Web Semântica para a integração de dados de execução financeira do Programa Dinheiro Direto na Escola (PPDE) e de indicadores educacionais, publicado em RDF (estrutura de triplas) para consumo das informações. O trabalho de Cabral et al. (2012) também apresenta um processo de integração de dados do Exame Nacional do Ensino Médio (Enem) realizado no ano de 2008. Apesar de utilizar de tecnologia de *Linked Data*, este trabalho não utiliza dados de outro domínio além da educação. Em d’Aquin et al. (2013), uma avaliação mostra como uma integração entre conjuntos de dados educacionais poderia melhorar a coesão da web com o alinhamento dos vocabulários. De modo geral, os trabalhos encontrados não abrangem em sua totalidade a integração de dados de educação com outros domínios, focando mais em análises com mineração de dados em um único domínio. O fato da pouca existência de casos de uso de dados abertos no Brasil deve ser mais uma motivação para estimular e disseminar a cultura de publicação de dados educacionais em formato aberto. O baixo uso pode influenciar uma futura relutância por parte de órgãos responsáveis pela publicação dando a impressão de baixa potencialidade de transparência e inovação (Alcantara et al. 2015).

3. Modelagem dos dados abertos conectados

3.1. Dados Abertos utilizados

Os dados referentes ao domínio da educação foram obtidos através do portal de Indicadores Educacionais do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP³) disponibilizados no formato “.xls”. Para esta pesquisa, foram utilizados apenas dados referentes ao ano de 2018:

- “**Quantidade Média de Alunos**” contém os dados pertinentes às escolas, como o código, nome, região, estado e a quantidade média de alunos por turma em cada nível escolar (1ª Série ao 3º Ano do ensino médio).
- “**Taxas de Rendimento**” engloba os dados pertinentes às escolas, como o código, o nome, região, estado e taxa de aprovação, reprovação e abandono em cada nível escolar (1ª Série ao 4º Ano do ensino médio).

Além de dados do domínio da educação, para escolas particulares, estaduais e públicas, foi utilizada uma base referente a dados do domínio de segurança para

³<http://portal.inep.gov.br/indicadores-educacionais>

obter os atendimentos da Guarda Municipal nos bairros das escolas. Essa base foi obtida através do portal de dados abertos da Prefeitura de Curitiba⁴ em formato “.csv”, referente ao ano de 2018:

- “**SigesGuarda**” contém os atendimentos feitos pela Guarda Municipal de Curitiba com o respectivo código do atendimento, tipo de atendimento e bairro de ocorrência.

Como as bases referentes aos dados da educação não disponibilizavam o dado referente ao bairro onde a escola está localizada, foi necessário utilizar uma quarta base referente ao planejamento de Curitiba que contém o registro de várias unidades públicas com seus respectivos bairros. Essa base foi obtida através do portal de dados abertos da Prefeitura de Curitiba³ em formato “.csv”:

- “**Unidades de Atendimento de Curitiba - Ativas**”, que pelo restante do artigo será chamado de “Unidades Ativas”. Essa base contém dados sobre unidades de atendimento de uso público disponíveis em Curitiba, como escolas, mercados, parques, entre outros. Os dados utilizados foram os pertinentes às escolas como: o nome e o código da escola, e o bairro em que se encontram.

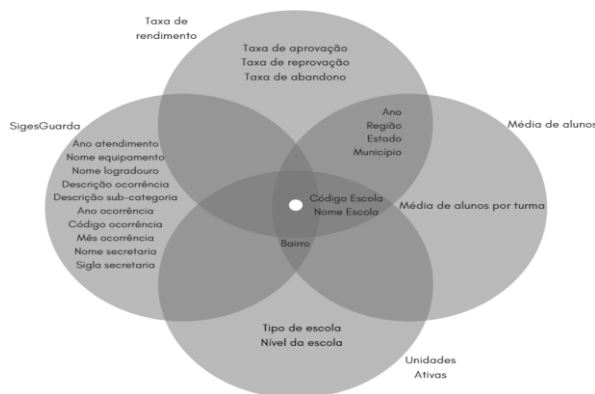


Figura 1. Dados utilizados na integração semântica de cada base

Antes de utilizar os dados foi necessário fazer um tratamento, como a redução da granularidade, para evitar inserir dados que não fizessem parte do escopo deste trabalho e para aumentar a acurácia dos links *owl:sameAs*. Os dados utilizados ao final do processo se referem a 8157 escolas, para 63 bairros da cidade de Curitiba. A Figura 1 mostra a complexidade de integrar dois domínios sociais diferentes, onde alguns dados são comuns a bases diferentes e mesmo assim a integração não é trivial pela questão semântica, em que um mesmo conceito pode estar representado por diferentes termos nas suas bases. Por exemplo, os dois dados chave para essa integração são o *código da escola* e o *bairro*, e não são dados presentes em todas as bases dos dois domínios e mesmo quando se encontram, não dividem da mesma homogeneidade semântica, apresentando termos diferentes para o mesmo conceito. Por exemplo, na base “SigesGuarda” o dado sobre *bairro* possui nome “ATENDIMENTO_BAIRRO_NOME” e na base “Unidades Ativas” o mesmo dado possui nome “NM_BAIRRO”. O mesmo ocorre com o código da escola, na base “Quantidade Média de Alunos” esse dado está com o nome “CO_ENTIDADE” e na base “Unidades Ativas” é “CD_LOCAL”.

3.2. Metodologia

⁴ <http://www.curitiba.pr.gov.br/dadosabertos/consulta>

Para a realização da integração foi utilizado como base o *framework* LDIF (*Linked Data Integration Framework*) (Schultz *et al.* 2012). Ele foi selecionado por se tratar de uma solução frequente na literatura para a integração de dados heterogêneos. Os passos para obter a visão LDM são: *i*) Extração dos dados abertos; *ii*) Materialização das triplas de acordo com ontologias locais; *iii*) Materialização de links *owl:sameAs* entre classes equivalentes; *iv*) Fusão dos links *owl:sameAs* e das triplas locais; *v*) Exportação do arquivo contendo todas as triplas para inserção em uma *Triple Store*. A Figura 2 apresenta as camadas resultantes dos passos do LDIF para a construção da visão LDM.

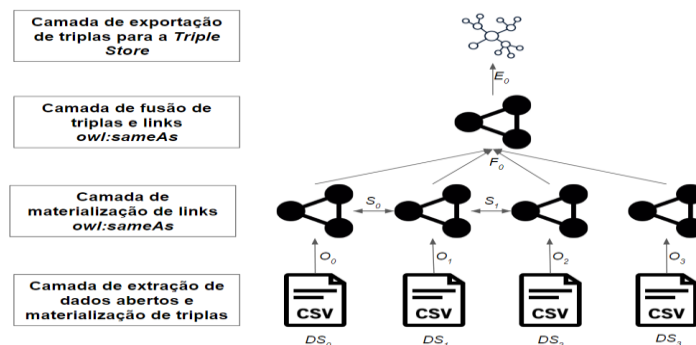


Figura 2. Visão de *Linked Data Mashup*

Na primeira camada da Figura 2 foi utilizada a ferramenta LOD-Refine⁵ para extrair e mapear os dados nos arquivos em “.csv” dos *datasets* $DS = \{DS_0, DS_1, DS_2, DS_3\}$, que são, respectivamente, “Taxa de Rendimento”, “Unidades Ativas”, “SigesGuarda” e “Quantidade Média de Alunos”, para as suas respectivas ontologias, gerando assim o conjunto $O = \{O_0, O_1, O_2, O_3\}$. Na segunda camada foi utilizada a ferramenta Silk (Volz *et al.* 2009) para a criação dos links *owl:sameAs*, resultando no conjunto de links $S = \{S_0, S_1\}$, onde S_0 representa a integração semântica pelo código das escolas nas triplas de O_0 (Taxa de Rendimento) e O_1 (Unidades Ativas) e S_1 representa a integração pelos bairros das triplas de O_1 (Unidades Ativas) e O_2 (SigesGuarda). Na terceira camada foi utilizada a ferramenta Sieve (Mendes *et al.* 2012) para fazer a fusão entre os arquivos das triplas das ontologias locais $\{O_0, O_1, O_2, O_3\}$, e os arquivos com os links *owl:sameAs* $\{S_0, S_1\}$, gerando um único arquivo F_0 e retirando a redundância nas triplas resultantes. Na quarta e última camada foi gerado no formato “*n-quad*” o arquivo final representado pelo conjunto E_0 na Figura 2, e inserido na *Triple Store* GraphDB⁶ para criar *queries* em SPARQL e extrair informações relevantes sobre os dois domínios integrados. Ao final do processo, para os dados utilizados e a ontologia construída, obteve-se um arquivo com 1.087.607 de triplas. Este arquivo de triplas final E_0 ⁷ está disponível na plataforma *osf.io* para ser acessado e consumido para extrair informações integradas.

4. Resultados e Discussões

Os resultados foram organizados para responder as perguntas de competência que nortearam a modelagem dos dados abertos conectados. Como o contexto social das escolas está sendo considerado para análise, foram criados dois grupos de observação: o

⁵ <https://github.com/sparkica/LODRefine>

⁶ <http://graphdb.ontotext.com/>

⁷ <https://osf.io/chygx/>

primeiro grupo contém os 5 bairros *mais inseguros* e o segundo grupo contém os 5 bairros *menos inseguros* da cidade de Curitiba. O critério de segurança considerado foi pela aplicação do filtro quanto ao “tipo de atendimento” feito pela Guarda Municipal, sendo selecionados os tipos: disparo de arma, lesão corporal, furto e roubo. Foram extraídos dados dos 3 anos do ensino médio referentes aos grupos de bairros, sendo mostrado aqui apenas os gráficos e discussões dos dados referentes ao 1º ano do ensino médio por limitação de espaço e porque o comportamento se repetiu no 2º e 3º ano do ensino médio. As queries em SPARQL que foram utilizadas para a extração dos dados integrados estão disponíveis⁶ juntamente com o arquivo gerado com as triplas integradas, bem como todos os gráficos gerados com os dados dos 3 anos do ensino médio.

Sobre a primeira questão de competência: *Existe uma relação entre o rendimento dos alunos e a quantidade de alunos por turma, com a quantidade de atendimentos feitos pela Guarda Municipal no bairro em que a escola se encontra?* Para endereçar esta discussão foram feitas consultas SPARQL⁶ para construir gráficos e também foi calculado coeficiente de correlação de Spearman. Os gráficos foram organizados em dois grupos: dados dos 5 bairros *mais inseguros* e dos 5 bairros *menos inseguros*, considerando a quantidade de atendimentos no bairro no ano de 2018. A Tabela 1 mostra os dados censitários do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) de 2010, especificamente a população total e a renda média, para os bairros selecionados. De forma geral, com exceções, os bairros mais seguros são os que possuem menores densidades populacionais e maiores rendas médias e os mais inseguros são os com maiores densidades populacionais e menores rendas médias.

Tabela 1. Dados sobre população e Renda Média dos bairros mais inseguros e menos inseguros.

Bairro	População	Renda Média (R\$)
Centro	37283	3347,35
Sítio Cercado	115525	1122,62
Cajuru	89784	793,75
Boqueirão	73178	1676,29
Cidade Industrial	172822	1146,67
Pilarzinho	28480	1827,46
Bairro Alto	46106	1672,46
Campo de Santana	7335	975
São Lourenço	6276	4169,23
Mercês	12907	3308,26

Na Figura 3 são explorados os dados do 1º ano nos 5 bairros *mais inseguros*. Dos 5 bairros considerados mais inseguros, o Centro é o único bairro que não é de periferia da cidade de Curitiba. Naturalmente as áreas centrais das grandes cidades tendem a possuir maior número de atendimentos pela Guarda Municipal, tanto pelo volume de pessoas que transitam, quanto pela facilidade de acesso pelo efetivo da Guarda Municipal. Já os outros 4 bairros possuem uma realidade social semelhante, tornando-os assim um cenário mais interessante de comparação. Desses 4 bairros, o considerado mais inseguro “Sítio Cercado” é o que possui a menor taxa de aprovação, indicando talvez uma relação entre a segurança e a educação. Também é notável que a taxa de aprovação nas escolas dos bairros *mais inseguros* (Figura 3.a) não passa de 80% (com exceção do Centro pelos motivos já apresentados), sendo que nos bairros *menos*

inseguros (Figura 3.b) apenas um bairro não apresenta taxa mínima de 80%. Na Figura 3.b são explorados os dados do 1º ano para os 5 bairros *menos inseguros* de acordo com o critério estabelecido neste trabalho. Especificamente para o filtro dos bairros menos inseguros também foi acrescentado outro critério que foi: a quantidade de registros da Guarda Municipal teria que ser maior do que 7, para evitar bairros muito isolados do centro, e que pelo menos 3 escolas estivessem presentes neste bairro para evitar a análise de bairros que possuíssem muito poucas ou nenhuma escola.

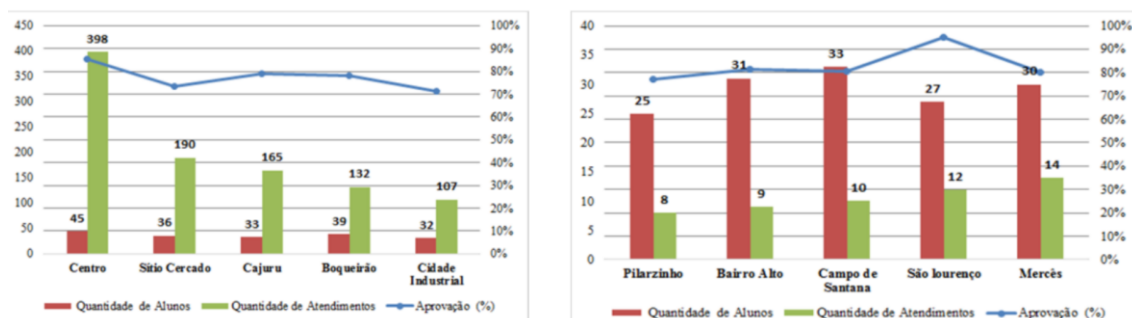


Figura 3. a) Média das taxas de aprovação e da quantidade de alunos do 1º Ano do Ensino Médio nos 5 bairros mais inseguros; b) Média das taxas de aprovação e da quantidade de alunos do 1º Ano do Ensino Médio nos 5 bairros menos inseguros.

Dos 5 bairros considerados menos inseguros, o Pilarzinho, Bairro Alto e Campo de Santana são considerados como periferia de Curitiba e possuem os índices de aprovação ainda menores se comparados com bairros São Lourenço e Mercês que são considerados bairros de condições sociais melhores em Curitiba. Essa visão possibilita perceber que não são somente os bairros com condições sociais melhores que são menos inseguros, porém ainda assim as escolas nesses bairros estão em desvantagens nos índices de aprovação. Na tentativa de utilizar o *mashup* criado não somente em discussões empíricas, foi também calculado o coeficiente de correlação de *Spearman*, adequado para situações em que a distribuição de probabilidade dos dados avaliados não é simétrica. Para este cálculo, a fim de obter um valor de correlação mais preciso (possível através de um tamanho de amostra maior), optou-se por utilizar como critério de bairros mais inseguros aqueles que tivessem uma quantidade de atendimentos maior que 10 (36 bairros), sendo os demais considerados como menos inseguros (26 bairros). Para o cálculo do coeficiente de correlação foram consideradas as variáveis “quantidade de atendimentos da guarda municipal no bairro” e “mediana das taxas de aprovação das escolas do bairro”. Os resultados obtidos apontam para correlações baixas e não significativas (nível de significância considerado foi 0,05): bairros mais inseguros ($r = 0.27$ e $p = 0.1$) e bairros menos inseguros ($r = 0.14$ e $p = 0.4$). Vários podem ser os fatores para este resultado como ausência de outros componentes sociais, tamanho de amostra ou até problema de qualidade dos dados abertos, e uma análise mais profunda e abrangente deve ser desenvolvida, porém, está além do escopo deste trabalho esta análise.

Sobre a segunda pergunta: *Existe uma diferença entre o rendimento e o tamanho das turmas entre escolas públicas e particulares de um bairro?* Para endereçar esta discussão foram feitas consultas SPARQL⁶ para construir gráficos. Na Figura 4 foram expostos os dados do 1º ano do ensino médio referentes a escolas públicas e particulares para os 3 bairros que possuem a maior densidade de escolas. É possível verificar um comportamento comum nos dados: onde a aprovação entre escolas públicas e particulares em média difere em quase 20 pontos percentuais, e que existe uma

tendência de escolas particulares possuírem turmas maiores. Além de reforçar uma das observações anteriores onde, com exceção do bairro Centro, o número de atendimentos parece ter uma relação com o rendimento dos alunos. Assim como na primeira pergunta essa observação foi feita de forma empírica, sem possibilidade de afirmar tais comportamentos sem o conhecimento de um profissional da área.

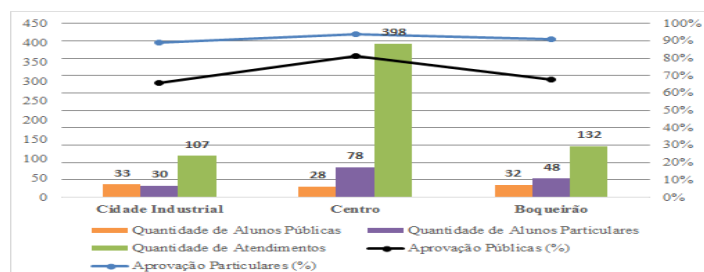


Figura 4. Média das taxas de aprovação e da quantidade de alunos do 1º Ano do Ensino Médio entre escolas particulares e públicas e a quantidade de atendimentos feitos no bairro.

Sobre a terceira pergunta: *Existe uma relação entre a taxa de abandono e o volume de atendimentos pela Guarda Municipal nos bairros das escolas?* Para endereçar esta discussão foram feitas consultas SPARQL⁶ para construir gráficos e também foi calculado índice de correlação de Spearman. Os gráficos também foram organizados em dois grupos: dados dos 5 bairros *mais inseguros* e dos 5 bairros *menos inseguros*, considerando a quantidade de atendimentos no bairro no ano de 2018. No gráfico da Figura 5, comparando as taxas de abandono em escolas presentes em bairros mais inseguros e menos inseguros, o que se nota é que na média, no primeiro ano do ensino médio o abandono é maior nos bairros mais inseguros; no segundo ano e terceiro ano a taxa de abandono é igual independentemente do nível de segurança.

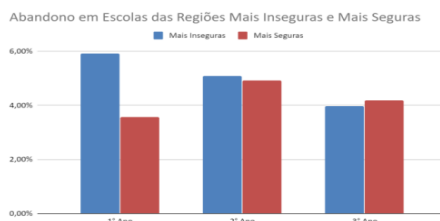


Figura 5. Média das taxas de abandono e da quantidade de alunos do 1º, 2º e 3º ano nas escolas dos bairros mais inseguros e mais seguros.

Da mesma forma que foi feito com a primeira pergunta, foi calculado o coeficiente de correlação de Spearman considerando o mesmo critério para classificar os bairros mais e menos inseguros. As variáveis utilizadas para calcular o coeficiente foram “quantidade de atendimentos da Guarda Municipal no bairro” e “mediana de taxas de abandono das escolas no bairro” para 63 bairros de Curitiba. Os resultados obtidos apontam para correlações baixas e não significativas: bairros mais inseguros ($r = -0.08$ e $p = 0.61$) e bairros menos inseguros ($r = -0.05$ e $p = 0.8$).

5. Considerações Finais

Esse artigo teve como objetivo mostrar como utilizar dados abertos conectados como apoio na discussão sobre a relação entre índices educacionais e outros domínios sociais. Para isso foram integradas bases de dados abertos dos domínios da segurança e educação da cidade de Curitiba, através do uso de ontologias juntamente do *framework*

LDIF. Através da criação do *mashup*, foi possível mostrar perguntas que podem ser extraídas de forma unificada sem a dificuldade de lidar com planilhas de dados abertos consideravelmente grandes e de manipulação não trivial pelo vocabulário de cada domínio, conforme exposto no corpo do estudo. As questões de competência formuladas mostraram a capacidade dos dados conectados responderem perguntas que combinavam realidades dos dois domínios, normalmente sendo necessário amplo tratamento de dados individuais. A riqueza de informações geradas pode auxiliar especialistas dos domínios interessados em visualizar realidades geralmente reportadas individualmente. Sobre a escalabilidade do processo, ela depende da disponibilidade de dados sobre educação e segurança de outras cidades e estados. Como a ontologia já está modelada, com o acesso a dados semanticamente semelhantes aos usados nesta pesquisa, é possível obter triplas de dados conectados para outros municípios e estados. Além de que a ontologia desenvolvida pode ser incorporada a outras ontologias para conectar e integrar dados de demais domínios presentes em centros urbanos com dados educacionais. Apesar de preliminares, os resultados obtidos a partir da manipulação do *mashup* tanto para construção de gráficos quanto para cálculo de correlação mostram as vantagens de se utilizar dados abertos no formato conectado para apoiar discussões em torno da educação. A conexão de novos dados, a revisão da qualidade dos dados abertos e o envolvimento de especialistas de domínio para interpretar as situações se fazem necessários para concretizar as vantagens do uso desses dados. Os dados integrados gerados e disponibilizados permitem a descoberta de novas informações através de consultas SPARQL e poderiam ser usados por especialistas de domínio, gestores e docentes para entender as relações sociais complexas que envolvem a educação e segurança pública das regiões em que eles atuam. Os resultados apresentados e os dados integrados podem ser usados para ações sociais, tomadas de decisão ou até mesmo para a criação de novas soluções baseadas em tecnologia da informação. A partir da compreensão da vantagem do uso de dados abertos conectados para auxiliar no processo de tomada de decisão, uma nova demanda surge que é reduzir a necessidade de conhecimento técnico em linguagem SPARQL para tornar as consultas mais próximas de especialistas de domínio, como gestores e tomadores de decisão. Como trabalhos futuros pretende-se conectar dados abertos de outros domínios e assim aumentar a visão integrada da educação e realizar análises junto a especialistas de domínio que possam oferecer novos *insights* sobre o resultado reportado até o momento e para novas discussões.

Referências

- Adeodato, P. J., Santos Filho, M. M., and Rodrigues, R. L. (2014). Predição de desempenho de escolas privadas usando o enem como indicador de qualidade escolar. In Simpósio Brasileiro de Informática na Educação -SBIE, volume 25, página 891.
- Alcantara, W., Bandeira, J., Barbosa, A., Lima, A., Ávila, T., Bittencourt, I., e Isotani, S. (2015). Desafios no uso de dados abertos conectados na educação brasileira. In DesafiE-4o Workshop de Desafios da Computação Aplicada à Educação. CSBC.
- Atenas, J., Havemann, L. (2015). “Open Data as Open Educational Resources: Case studies of emerging practice”. London: Open Knowledge, Open Education Working Group.
- Avila, C. V. S. ; Rolim, T. V. ; Cruz, M. M. L. ; Silva, J. W. F. ; Venceslau, A. D. P. ; Vidal, V. M. P. (2018) Um Linked Data Mashup de Dados de Execuções Financeiras e Indicadores Educacionais no Ensino Básico. In: XXIX Simpósio Brasileiro de Informática em Educação (SBIE), 2018, Fortaleza - Ceará. Anais do SBIE 2018 v. 29.

- Bandeira, J., Ávila, T., Alcantara, W., Sobrinho, A., Bittencourt, I. I., and Isotani, S. (2015). Dados abertos conectados para a educação. *Jornada de Atualização em Informática na Educação*, 4(1):47–69.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific american*, 284(5):34–43.
- Bizer, C. and Seaborne, A. (2004). D2rq-treating non-rdf databases as virtual rdf graphs. In *Proceedings of the 3rd international semantic web conference (ISWC2004)*, volume 2004. *Proceedings of ISWC2004*
- Cabral, S. P., Beduschi, N. B., Zancanaro, A., Todesco, J. L., and Gauthier, F. A. O. (2012). Aplicando linked data na publicação de dados do Enem. In *ONTOBRAS- MOST*, pages 176–181.
- Carvalho, J., Cruz, L., and Gouveia, R. (2017). Descoberta de conhecimento com aprendizado de máquina supervisionado em dados abertos dos censos da educação básica e superior. In *Anais dos Workshops do CBIE*, volume 6, página 674.
- d’Aquin, M., Adamou, A., e Dietze, S. (2013). Assessing the educational linked data landscape. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 43–46. ACM.
- Ferreira, G. (2015). Investigação acerca dos fatores determinantes para a conclusão do ensino fundamental utilizando mineração de dados educacionais no censo escolar da educação básica do inep 2014. In *Anais dos Workshops do CBIE*, volume 4, página 1034.
- Isotani, S., Bittencourt, I. I. (2015). “Dados abertos conectados”. Editora Novatec.
- Mendes, P. N., Mühleisen, H., and Bizer, C. (2012). Sieve: linked data quality assessment and fusion. In *Proceedings of the Joint EDBT/ICDT Workshops*, pages 116–123. ACM.
- Penteado, B., Bittencourt, I. I., Isotani, S. (2017). Dados abertos educacionais no Brasil e sua preparação para os dados abertos na web. In: *XXVIII Simpósio Brasileiro de Informática na Educação SBIE (Brazilian Symposium on Computers in Education)*, 2017, Recife. p. 526.
- Rigo, S. J., Cazella, S. C., and Cambrozzi, W. (2012). Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In *Anais do Congresso Brasileiro de Informática na Educação (CBIE)*, páginas 168–177.
- Santos, P., Ferreira, R., Miranda, P. (2017) Dados Abertos Educacionais: Uma Revisão da Literatura Brasileira. In *Simpósio Brasileiro de Informática na Educação -SBIE*, volume 28, página 11.
- Sarker, Farhana (2014) “Linked data technologies to support higher education challenges: student retention, progression and completion”. University of Southampton, Physical Sciences and Engineering. Tese de doutorado, 273pp.
- Schultz, A., Matteini, A., Isele, R., Mendes, P. N., Bizer, C., and Becker, C. (2012). LDIF - A Framework for Large-Scale Linked Data Integration. In *21st WWW, Developers Track*, page to appear.
- Pino, A. (2007). Violência, educação e sociedade: um olhar sobre o Brasil contemporâneo. *Educação & sociedade*, páginas 763-785.
- Volz, J., Bizer, C., Gaedke, M., and Kobilarov, G. (2009). Silk-a link discovery framework for the web of data. *LDOW*, 538.
- W3C (2017) “Data on the Web Best Practices”. W3C Recommendation 31 January 2017.