

# Web Scraping em Marketplace para detecção do uso indevido de Vale-Transporte

João Carlos dos Santos Rodrigues da Silva<sup>1</sup>; João Pedro Fagundes Calazans<sup>2</sup>; Lucas Batista de Azevedo<sup>3</sup>.

<sup>1</sup>Semove, Rua da Assembléia, nº 10 – 39º andar – Centro – Rio de Janeiro/RJ – CEP: 20.011-901, (21) 3221-6300, e-mail: [joao.silva@semove.org.br](mailto:joao.silva@semove.org.br);

<sup>2</sup>Semove, Rua da Assembléia, nº 10 – 39º andar – Centro – Rio de Janeiro/RJ – CEP: 20.011-901 (21) 3221-6300, e-mail: [joao.calazans@semove.org.br](mailto:joao.calazans@semove.org.br);

<sup>3</sup>Semove, Rua da Assembléia, nº 10 – 39º andar – Centro – Rio de Janeiro/RJ – CEP: 20.011-901 (21) 3221-6300, e-mail: [lucas.azevedo@semove.org.br](mailto:lucas.azevedo@semove.org.br).

## RESENHA

Neste artigo é apresentado uma análise de fraudes envolvendo anúncios de venda de vale-transporte no Estado do Rio de Janeiro em um site de *marketplace*. Para a coleta de dados, foi aplicado o método de *web scraping* para identificar anúncios com termos que possibilitam uma investigação a fim de verificar se o anúncio se trata de uma fraude com vale-transporte. Após a consolidação de uma relação de anúncios validados, é realizado uma denúncia na própria plataforma.

PALAVRAS-CHAVES: Vale-Transporte, Transporte público, Fraude, Web Scraping

## INTRODUÇÃO

O vale-transporte surgiu como uma concessão voluntária aos funcionários no Brasil, em 1985, com a lei nº 7.418 (BRASIL, 1985). Porém, a partir de 1987, através da Lei nº 7.619 (BRASIL, 1987), o benefício do vale-transporte passou a ser uma obrigação das empresas, conforme o artigo 4º da lei: “A concessão do benefício ora instituído implica a aquisição pelo empregador dos Vales-Transporte necessários aos deslocamentos do trabalhador no percurso residência-trabalho e vice-versa, no serviço de transporte que melhor se adequar.” Ou seja, o benefício garante ao trabalhador a mobilidade, fator esse que é definida por Silva (SILVA, 2013) como “a capacidade de deslocamento de pessoas e bens de um ponto de origem a outro ponto de destino dentro do espaço para o desenvolvimento de suas atividades em tempo e modo adequados e de forma segura”.

No Estado do Rio de Janeiro esse benefício é fornecido por meio de cartão eletrônico desde 2004, após a vigência da Lei nº 4291 (ALERJ,2004) que “institui o sistema de bilhetagem eletrônica nos serviços de transporte coletivo de passageiros por ônibus, de competência do estado do rio de janeiro e dá outras providências.” substituindo o vale de papel por cartão vale-transporte conforme o artigo 3º da lei: “O vale-transporte será obrigatoriamente emitido sob a forma de cartão eletrônico, observado o Parágrafo único, do art. 4º e respeitada a legislação federal, possibilitando a sua utilização em outros tipos de serviços de interesse das delegatárias dos serviços de transporte rodoviário intermunicipal de passageiros.”

Todavia, nem sempre o cartão é utilizado para seu fim, sendo comercializado para terceiros, onde é caracterizado como fraude. Fraude é definido (FERREIRA, 2010) como

qualquer ato ou omissão intencional concebido para enganar os outros, resultando na vítima sofrendo perdas e/ou o autor obtendo um ganho. A prática da fraude promove impactos negativos para o empregador e ao operador da bilhetagem eletrônica, podendo se estender ao poder público e sociedade como um todo.

A empresa é afetada em virtude do desvio da finalidade no pagamento do vale-transporte dos funcionários, podendo elevar os custos da empresa. Já o operador da bilhetagem tem um impacto direto no planejamento de transporte, visto que pode distorcer as estatísticas de demanda de transporte e afetar o planejamento de rotas e serviços. Se os dados de uso do transporte público forem inflados devido à fraude, isso pode levar a decisões equivocadas sobre a oferta de transporte em determinadas áreas, resultando em ineficiências operacionais, além de prejudicar o fluxo de caixa da empresa de bilhetagem eletrônica.

Na esfera pública, esse tipo de fraude pode gerar subsídios indevidos, no caso do Estado do Rio de Janeiro é evidenciado pela utilização inadequada do Bilhete Único Intermunicipal – BUI que é um benefício social aplicado nas tarifas de transporte público e concedido pelo Governo do Estado do Rio de Janeiro. O passageiro cadastrado no programa BUI pode utilizar até 2 (dois) meios de transporte, um deles obrigatoriamente intermunicipal, no período de 3 (três) horas, pagando o valor máximo de R\$ 8,55 (oito reais e cinquenta e cinco centavos). A diferença entre o valor máximo pago pelo passageiro (de R\$ 8,55) e o custo real dos transportes escolhidos no deslocamento realizado é complementado pelo Governo do Estado. Tal benefício é utilizado por passageiros que utilizam cartões vale-transporte de cambistas, que em uma negociação fraudulenta utilizam dois meios de transporte pagando ao máximo R\$8,55, que além de lucrar com a venda de duas passagens ultrapassando R\$8,55, ainda gera um subsídio incorreto ao Governo do Estado do Rio de Janeiro.

A comercialização do cartão nem sempre é feita de forma física, sendo utilizado sites de *marketplace*, que reúne variados vendedores e lojas, com diferentes ofertas e diferentes produtos disponíveis. Estas lojas virtuais podem estar em redes sociais que abrem esse espaço para os vendedores, como por exemplo o *Facebook* que destina uma parte da sua plataforma para que os usuários vendam seus produtos.

## WEB SCRAPING

Diante do contexto da comercialização de vale-transporte, ao investigar a presença de anúncios em diferentes plataformas, notou-se um quantitativo discrepante de divulgações no *Facebook* em relação aos demais. Para identificação e combate do exposto, foi utilizado a técnica denominada *web scraping*.

A técnica *web scraping*, ou “raspagem de dados”, pode ser definida como um processo de extração utilizado para coletar dados relevantes na *web* de forma sistemática (LOURENÇO et al., 2013). Calò (CALÒ, 2014) destaca a importância das ferramentas capazes de extrair o vasto volume de dados que permeia as redes, visando obter uma análise mais precisa. Ele salienta que o progresso da computação nos últimos anos tem impulsionado a geração massiva de dados em todas as áreas do conhecimento. Como resposta, surgiram novas abordagens de análise cujo principal propósito é extrair informações relevantes a partir dos extensos conjuntos de dados brutos.

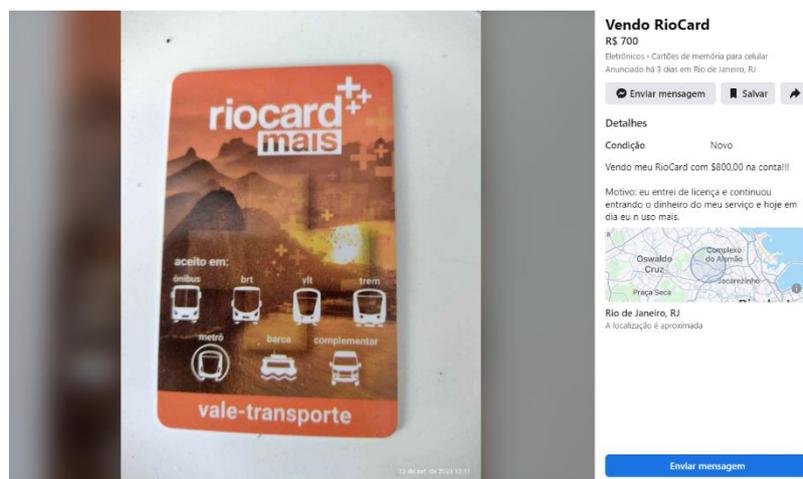
Segundo Vlad Krotov e colaboradores (KROTOV; TENNYSON, 2021), o *web scraping* é composto por três fases: análise da página *web*; rastreamento da página *web*; e organização dos dados. Estas fases necessitam de alguns requisitos, a primeira requer entendimento de algumas tecnologias de *web* e conhecimento básico da estrutura do *World Wide Web*, já a

segunda fase requer o entendimento de pelo menos uma linguagem de programação popular, como R ou *Python*.

A fim de utilizar essa abordagem, é necessário desenvolver um script por meio da programação que tenha a capacidade de extrair informações de uma ampla quantidade de dados dispersos em várias páginas e guias de um site específico de forma automática. Entretanto, é crucial incorporar expressões regulares no script para identificar um padrão no código-fonte da página *web* em análise, permitindo assim a extração dos dados pertinentes do site. Após a conclusão da extração, é viável armazenar os dados em tabelas para simplificar o processo de análise.

## DIAGNÓSTICO, PROPOSIÇÕES E RESULTADOS

Primeiramente, investigou-se a presença de anúncios de venda e aluguel de cartões de vale-transporte em sites de *Marketplace* e *E-Commerce*. Após essa etapa, notou-se que os anúncios se encontram majoritariamente na plataforma *Facebook Marketplace*, o que levou a motivação de focar nessa plataforma para o desenvolvimento do algoritmo. A figura 1 mostra um caso típico de venda de vale-transporte no *Facebook Marketplace*.



Vendo meu RioCard com \$800,00 na conta!!!

Motivo: eu entrei de licença e continuou entrando o dinheiro do meu serviço e hoje em dia eu n uso mais.

**Figura 1** – Exemplo de Anúncio de Venda de vale-transporte no *Facebook Marketplace*

Após decidir qual *website* seria o foco, precisou-se analisar o código fonte da página *web* do *Facebook Marketplace* para que em seguida o algoritmo fosse desenvolvido com base nos elementos do site. Para a programação do algoritmo foi utilizado a linguagem de programação *Python*, e internamente foram aplicadas as bibliotecas *Beautiful Soup* e *Selenium*. Além do *web scraping*, também foi utilizado o *Microsoft Excel* para análise descritiva e elaboração dos gráficos.

A figura 2 demonstra de forma macro o processo de desenvolvimento do algoritmo na presente pesquisa.

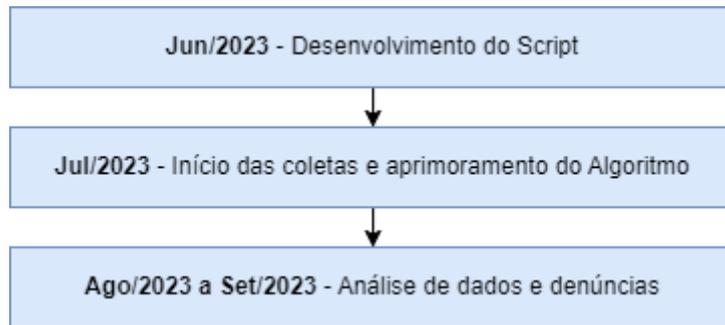


Figura 2 – Cronograma de Atividades

Primeiramente são fixadas as configurações iniciais para a inicialização do algoritmo, tais como: configurações do *driver* do navegador *Google Chrome* e credenciais do *Facebook* para Login. Posteriormente, são definidas as *urls* com palavras chaves no *Facebook Marketplace* que o algoritmo irá percorrer para identificar os anúncios de vendas e aluguéis dos cartões.

O *Selenium* atrelado ao *Chrome Web Driver* simula um usuário utilizando o navegador, onde cada *url* predefinida é acessada de forma automática. A biblioteca *Beautiful Soup* é responsável por fornecer mecanismos que possibilitam a extração do código fonte de cada uma dessas páginas e busca os elementos de interesse, que são: Título do anúncio, região do anúncio, saldo do cartão, preço da venda, *link* da página *web*, tipo do cartão e código do produto. A figura 3 mostra o algoritmo em funcionamento.

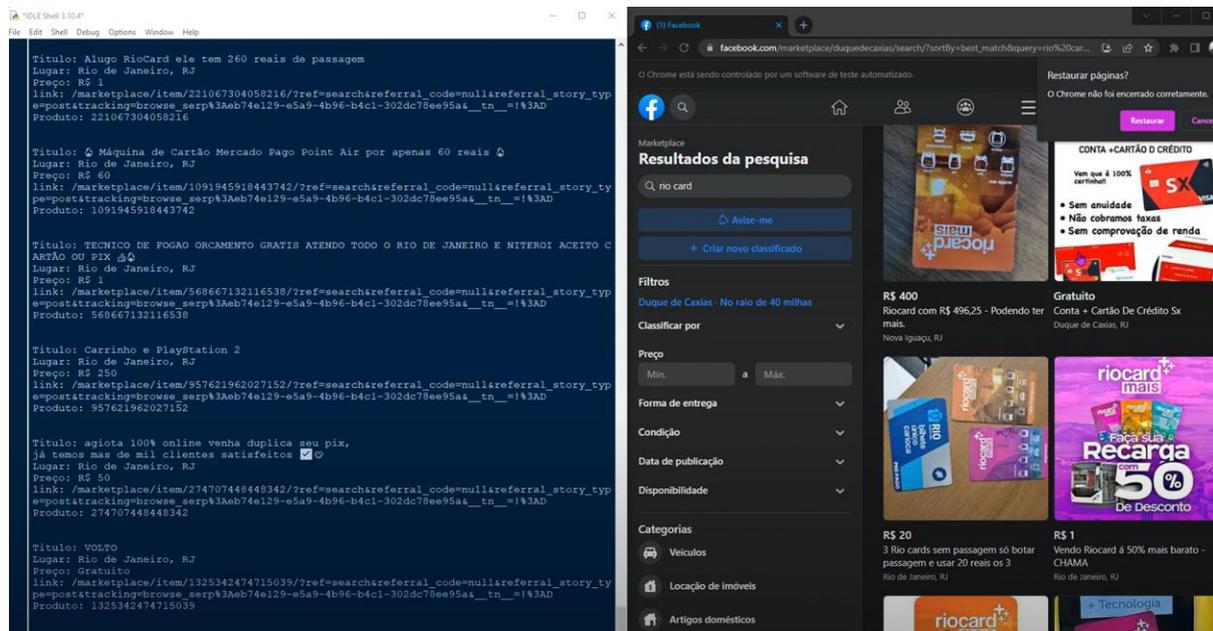


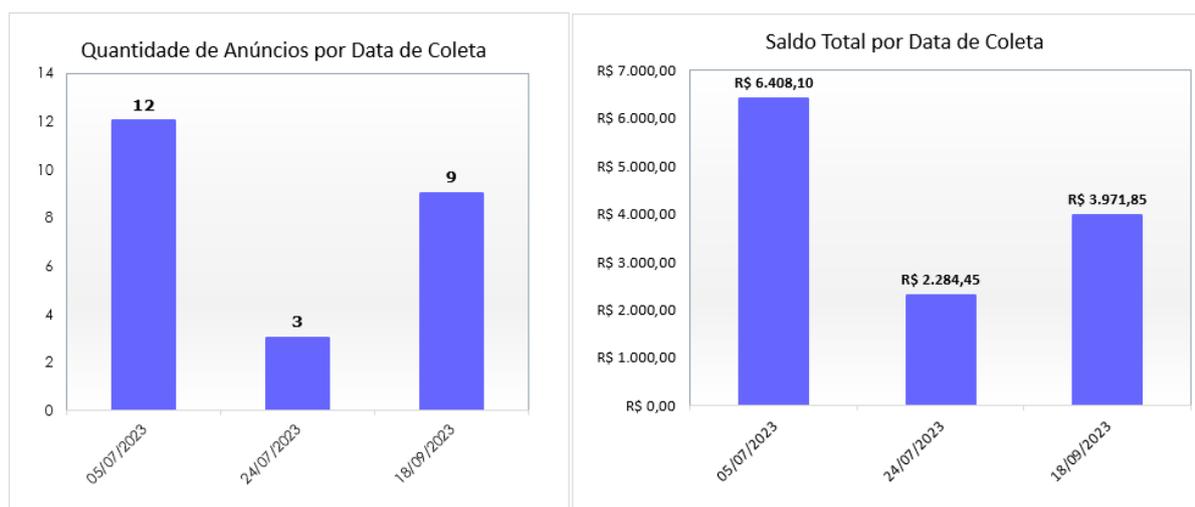
Figura 3 – Funcionamento do Algoritmo

Ao passo que esses dados são coletados de cada página, é feito um armazenamento em um *dataframe*, que é uma estrutura de dados similar a uma tabela. A tabela criada passa por uma análise de acordo com algumas palavras chaves, no intuito de confirmar se realmente aquele anúncio interessa ao estudo.

**Tabela 1 – Amostra de Anúncios (Exemplo)**

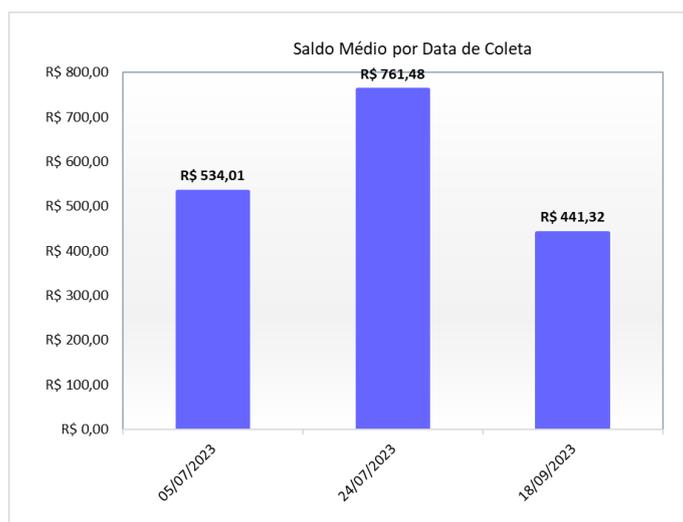
Data da Coleta	Saldo	Preço	Região	Link	Cartão	Código do Produto
05/07/2023	R\$ 700,00	R\$ 350,00	Mesquita	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	648873550445046
05/07/2023	R\$ 1.304,00	R\$ 650,00	Itaguaí	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	1936071946753990
05/07/2023	R\$ 1.280,00	R\$ 500,00	Rio de Janeiro	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	219551270679812
05/07/2023	R\$ 540,00	R\$ 300,00	Rio de Janeiro	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	6784642891549360
24/07/2023	R\$ 433,75	R\$ 300,00	São João de Meriti	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	1384148965503330
18/09/2023	R\$ 590,00	R\$ 400,00	Nova Iguaçu	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	1274666393249710
18/09/2023	R\$ 800,00	R\$ 700,00	Rio de Janeiro	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	325673109934237
18/09/2023	R\$ 800,00	R\$ 800,00	Duque de Caxias	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	976345340137397
18/09/2023	R\$ 800,00	R\$ 600,00	Campos dos Goytacazes	<a href="https://www.facebook.com/marketplace/iten">https://www.facebook.com/marketplace/iten</a>	VT	1288572271826620

Ao todo, foram realizadas três coletas de dados, nos meses de julho e setembro de 2023, totalizando 24 anúncios espalhados em diversos municípios do estado do Rio de Janeiro. O Gráfico 1 exibe a quantidade de anúncios e saldo total por data de coleta.



**Gráfico 1 – Quantidade de Anúncios e Saldo Total por Data de Coleta**

O saldo e quantidade de anúncios são proporcionais em cada data de coleta, havendo um destaque para o dia 05/07/2023, com um total de 12 anúncios de comercialização de vale-transporte por terceiros, totalizando um saldo R\$6.408,10 nestes anúncios. Porém, ao analisar o saldo médio por data de coleta, o dia 24/07/2023 teve um maior destaque, conforme exposto no Gráfico 2.



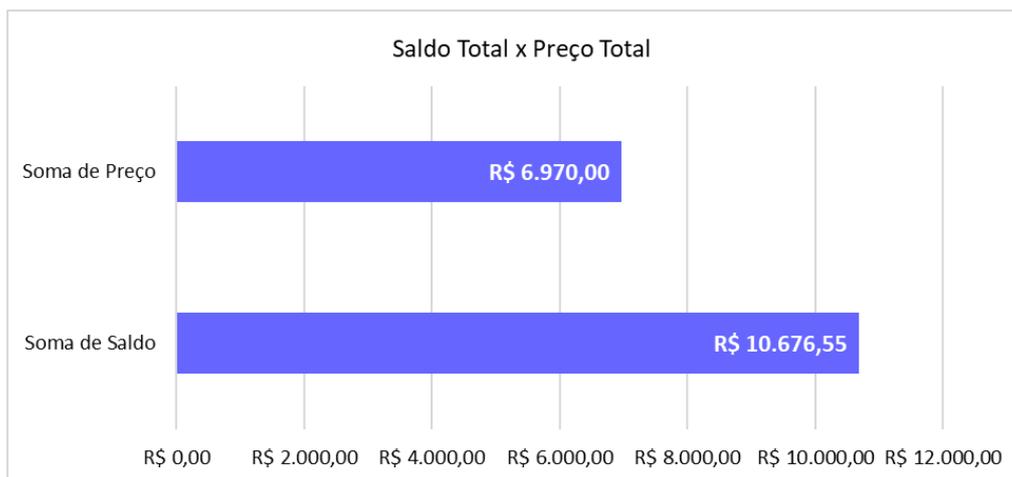
**Gráfico 2 – Saldo Médio por Data de Coleta**

Ao analisar a quantidade de anúncios por município, foi observado que a maior parte se encontra dentro da Região Metropolitana do Estado do Rio de Janeiro, representando 92% do todo. O Gráfico 3 exibe a quantidade de anúncios por município.



**Gráfico 3 – Quantidade de Anúncios por Município**

Ao verificar cada anúncio individualmente, foi percebido uma diferença entre o preço que o portador do cartão vale-transporte cobra na plataforma e o saldo presente nesse cartão. Dentre os 24 anúncios analisados, foram considerados os casos onde o anunciante deixa explícito o preço, para que fosse possível quantificar em média a diferença entre o saldo do cartão e o preço do anúncio. Foi observado que a diferença média é de 28,4% a menos para o preço do anúncio. No intuito de mensurar valores absolutos, notou-se uma diferença de 34,7% entre o total de saldo nestes cartões e o total de preço cobrado nos anúncios. Os valores absolutos podem ser verificados no Gráfico 4.



**Gráfico 4 – Saldo Total x Preço Total**

Após a análise e validação dos anúncios, foram feitas denúncias através da própria plataforma do *Facebook Marketplace*. Esse mecanismo permite que o próprio *Facebook* analise caso a caso e remova o anúncio de acordo com seu julgamento.

## CONSIDERAÇÕES FINAIS

A pesquisa foi importante para mensurar um tipo de fraude que surgiu com o advento do uso de cartões eletrônicos como vale-transporte e o alcance gerado pelas plataformas de *marketplace*. A linguagem *Python* é gratuita, bem como as bibliotecas utilizadas para a aplicação da técnica de *web scraping*, logo, não houve custo algum para esse tipo de análise.

A área de estudo desta pesquisa foi o Estado do Rio de Janeiro, porém existe a possibilidade desta ser aplicada em outros estados do país, auxiliando assim os operadores de bilhetagem eletrônica e o poder público.

Recomenda-se para trabalhos futuros o aprimoramento do algoritmo para realizar coletas diárias, sendo ele hospedado em um ambiente na nuvem utilizando tecnologias como o *Google Cloud*.

Por fim, além das denúncias, seria interessante uma parceria entre as operadoras de bilhetagem eletrônica e as empresas de *marketplace*, com o intuito de reportar esses casos de venda de vale-transporte diretamente.

## REFERÊNCIAS

ALERJ. Lei no 4.291, de 22 de março de 2004. Institui a bilhetagem eletrônica e substitui o vale de papel por cartão vale-transporte. Assembleia Legislativa do Estado do Rio de Janeiro, Rio de Janeiro, RJ, 2004.

BRASIL. Lei no 7.418, de 16 de dezembro de 1985. Institui o vale-transporte, que o empregador, pessoa física ou jurídica, antecipa ao empregado para utilização efetiva em despesas de deslocamento residência-trabalho e vice-versa. *Diário Oficial da República*

*Federativa do Brasil*, Brasília, DF, 1985. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/LEIS/L7418.htm](http://www.planalto.gov.br/ccivil_03/LEIS/L7418.htm)>. Acesso em: 16 set. 2023.

BRASIL. Lei no 7.619, de 30 de setembro de 1987. altera a lei no 7.418 e torna obrigatório o benefício do vale-transporte. *Diário Oficial da República Federativa do Brasil*, Brasília, DF, 1987. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/leis/l7619.htm](https://www.planalto.gov.br/ccivil_03/leis/l7619.htm)>. Acesso em: 16 set. 2023.

CALÒ, A. *Extração e análise de informações jurídicas públicas*. Tese (Graduação em Ciência da computação) — Universidade de São Paulo (USP), São Paulo, 2014. Disponível em: <<https://bcc.ime.usp.br/tccs/2014/sandro/Monografia.pdf>>. Acesso em: 16 set. 2023.

FERREIRA, A. B. de H. *Dicionário Aurélio da língua portuguesa*. [S.l.]: Editora Positivo, 2010.

KROTOV, V.; TENNYSON, M. *Web Scraping in the R Language: A Tutorial*. *Journal of the Midwest Association for Information Systems (JMWAIIS)*, v. 2021, n. 1, p. 61–77, 01 2021. Disponível em: <<https://aisel.aisnet.org/jmwais/vol2021/iss1/5>>. Acesso em: 16 set. 2023.

LOURENÇO, A. et al. *Web scraping technologies in an API world*. *Briefings in Bioinformatics*, v. 15, n. 5, p. 788–797, 04 2013. ISSN 1467-5463. Disponível em: <<https://doi.org/10.1093/bib/bbt026>>. Acesso em: 16 set. 2023.

SILVA, F. N. da. *Mobilidade urbana: os desafios do futuro*. *Cadernos Métopole*, v. 15, n. 30, p. 377–388, 2013.