

Uma Metodologia para Coleta e Mapeamento de Dados de Licitações Públicas dos Portais da Transparência dos Municípios de Santa Catarina

Breno da Silva Pereira¹, Carina F. Dorneles¹

¹Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brazil

brenosilvp@gmail.com, carina.dorneles@ufsc.br

Abstract. *Public transparency is a fundamental pillar of good governance, enabling citizens and oversight bodies to monitor the use of public resources. However, the way procurement data is made available on municipal transparency portals often presents challenges, such as a lack of standardization, difficulty of access, and the need for manual data extraction. These obstacles hinder the analysis and monitoring of public contracts. Given this scenario, this work proposes an automated approach for collecting and mapping procurement data from the transparency portals of municipalities in Santa Catarina. The solution aims to centralize and structure the information provided by local governments, making data access and analysis more efficient. The study presents the problem's context, discusses the motivation behind the proposed solution, details the methodology used, and explores the results obtained so far.*

Resumo. *A transparência pública é um pilar fundamental para a boa governança, permitindo que cidadãos e órgãos de controle acompanhem o uso de recursos públicos. No entanto, a forma como os dados de licitações são disponibilizados nos portais da transparência de municípios muitas vezes apresenta desafios, como a falta de padronização, dificuldade de acesso e a necessidade de extração manual de informações. Esses obstáculos comprometem a análise e o monitoramento das contratações públicas. Diante desse cenário, este trabalho propõe uma abordagem automatizada para a coleta e o mapeamento de dados de licitações nos portais da transparência dos municípios de Santa Catarina. A solução busca centralizar e estruturar as informações disponibilizadas pelas prefeituras, facilitando o acesso e a análise dos dados. O estudo apresenta o contexto do problema, discute a motivação para o desenvolvimento da solução, detalha a metodologia empregada e explora os resultados obtidos até o momento.*

1. Introdução

O Portal da Transparência, de acordo com a Controladoria-Geral da União (CGU), é uma ferramenta essencial para o acompanhamento do uso dos recursos públicos, permitindo maior auditabilidade e controle social¹. Ele possibilita que cidadãos, pesquisadores

¹<https://portaldatransparencia.gov.br/ontrole-social/o-portal-como-ferramenta>

e órgãos fiscalizadores acessem informações detalhadas sobre despesas governamentais, como contratos e investimentos em diversas áreas. Ao disponibilizar dados de forma acessível e atualizada, os portais da transparência promovem participação popular, ajudam a prevenir irregularidades e garantem maior eficiência na gestão pública. Nesse sentido, dados referentes a licitações tornam-se fundamentais para garantir o correto acompanhamento do uso das verbas públicas. De acordo com [de Mello 2004], “Licitação é o procedimento administrativo pelo qual uma pessoa governamental, pretendendo alienar, adquirir ou locar bens, realizar obras ou serviços, segundo condições por ela estipuladas previamente, convoca interessados na apresentação de propostas, a fim de selecionar a que se revele mais conveniente em função de parâmetros antecipadamente estabelecidos e divulgados”.

Santa Catarina possui 295 municípios, e cada um deles possui seu próprio portal da transparência, nos quais são disponibilizadas informações sobre receitas, despesas, contratos, licitações e demais dados relacionados à administração pública local. No entanto, esses portais têm uma estrutura específica, conteúdos que não são padronizados para os 295 municípios, e formas de acesso muitas vezes dificultadas, como o uso de captcha para acesso às informações. Essa sistemática de disponibilização de dados dificulta a visão ampla e integrada dos gastos públicos no estado. Para cidadãos, pesquisadores e órgãos fiscalizadores, essa heterogeneidade e especificidade de disponibilização de dados prejudica a tarefa de comparar dados entre diferentes municípios, exigindo consultas individuais em diversas plataformas.

O objetivo deste trabalho é desenvolver e propor uma abordagem para a coleta e mapeamento de dados relacionados a licitações disponibilizados nos portais da transparência dos municípios de Santa Catarina. Essa solução visa superar os desafios de acesso descentralizado e dispersão das informações, oferecendo uma forma centralizada e eficiente de acesso aos documentos e dados públicos fornecidos pelas prefeituras. Com isso, busca-se consolidar, em um formato estruturado, as informações abertas referentes a processos licitatórios, promovendo maior acessibilidade, organização e potencial para análises comparativas e estudos aprofundados sobre os gastos públicos no estado.

Este documento está estruturado da seguinte maneira: na Seção 2 são detalhados os trabalhos relacionados; a Seção 3 apresenta a metodologia utilizada; a Seção 4 descreve o processo de desenvolvimento do trabalho, enquanto a Seção 5 exibe uma discussão a respeito dos resultados obtidos; finalmente, a Seção 6 conclui o artigo e oferece possíveis caminhos para trabalhos futuros.

2. Trabalhos Relacionados

Nesta Seção, a análise de trabalhos relacionados é apresentada em duas partes principais: (i) trabalhos da literatura científica: que apresentam pesquisas e estudos teóricos sobre o tema, onde são apresentadas as diferentes abordagens, metodologias e técnicas utilizadas; e (ii) ferramentas desenvolvidas: que descrevem softwares e sistemas práticos para realizar tarefas de extração e processamento de documentos, onde são apresentadas as funcionalidades, características e aplicações de cada ferramenta, destacando seus pontos fortes e fracos em relação ao tema em questão.

Alguns trabalhos da literatura exploram técnicas para automatizar a coleta e análise de dados a partir de portais governamentais da transparência.

Os trabalhos de [de Moraes Almeida and Silva 2018], [de Assis and Gomide 2021] e [Galdino et al. 2020] destacam-se por suas contribuições na automação da coleta e análise de dados de transparência pública em portais municipais. O sistema proposto por [de Moraes Almeida and Silva 2018] se sobressai pela capacidade de identificar inconformidades e erros de categorização nos dados de licitações, enquanto o estudo de [de Assis and Gomide 2021] demonstra a eficácia do web scraping ao transformar dados desestruturados em informações acessíveis, além de desenvolver um chatbot interativo para facilitar o acesso às informações. Já a aplicação de [Galdino et al. 2020] se destaca pela disponibilização dos dados em dashboards interativos, promovendo um controle social mais eficiente. No entanto, um ponto fraco comum a todos esses trabalhos, em comparação com a metodologia proposta neste artigo, é a limitação ao foco em portais específicos de um único município, o que restringe a abrangência das análises e não considera os desafios de padronização enfrentados em um contexto mais amplo, como o dos 295 municípios de Santa Catarina.

Com a disponibilidade dos dados através dos portais de transparência, muitas ferramentas surgiram com o intuito de coletar os dados e torná-los úteis para processamento. As ferramentas desenvolvidas `crawler-portal-transparencia`², `crawler-transparencia`³, `crawler`⁴, `portal_transparencia_am`⁵ e `transparency-scrapper`⁶ apresentam contribuições relevantes para a automação da coleta de dados de transparência pública. A solução `crawler-portal-transparencia` se destaca pela escalabilidade proporcionada pelo uso de MongoDB e uma API RESTful para consultas eficientes, mas sua dependência da estrutura fixa do Portal da Transparência do Governo Federal compromete sua adaptabilidade. A aplicação `crawler-transparencia` permite personalizar consultas detalhadas sobre despesas públicas, embora sua limitação a períodos de um mês por extração reduza sua eficiência em análises de longo prazo. Já o `crawler` automatiza com eficiência a coleta periódica e facilita a execução com Docker, mas é restrito ao portal específico de João Pessoa, dificultando sua aplicação em outros contextos. O `portal_transparencia_am` se destaca pela capacidade de consolidar os arquivos CSV extraídos, porém sua funcionalidade se restringe a dados salariais do Amazonas, sem flexibilidade para outros tipos de informação. O `transparency-scrapper`, por sua vez, possibilita a extração automatizada de dados do Portal da Transparência de Teutônia/RS, mas sua abrangência é limitada a um único município, restringindo sua aplicabilidade em contextos mais amplos. Comparadas à metodologia proposta neste artigo, todas as ferramentas são limitadas por seu foco em portais únicos, enquanto a abordagem desenvolvida aqui se diferencia por abranger os 295 municípios de Santa Catarina, ampliando a capacidade de análise em um cenário mais complexo e diverso.

3. Metodologia LicitScraper-SC

A proposta de extração automatizada de dados de licitações dos municípios de Santa Catarina pode ser descrita em três etapas, a saber: (i) mapeamento das URLs dos portais

²<https://github.com/marck010/crawler-portal-transparencia>

³<https://github.com/LucasBassetti/crawler-transparencia>

⁴<https://github.com/minha-cidade/crawler>

⁵https://github.com/julianyraiol/portal_transparencia_am

⁶<https://github.com/joelsouza/transparency-scrapper>

da transparência dos municípios de Santa Catarina; (ii) obtenção dos links para dados disponíveis nos portais; e (iii) extração de dados e documentos referentes às licitações nos portais da transparência. A Figura 1 mostra a proposta de desenvolvimento, representando numericamente as etapas informadas, que são descritas nas seções a seguir.

- I. Primeiramente, desenvolver uma pesquisa automatizada em uma ferramenta de busca, utilizando a palavra “prefeitura” e o nome do município como referências, a fim de armazenar as URLs dos portais das prefeituras municipais. Posteriormente, acessar as URLs conseguidas e extrair os links para as transparências.
- II. A partir das URLs dos portais da transparência, produzir um extrator para conseguir os links dos dados disponibilizados por cada um deles.
- III. Em posse das URLs dos dados de todos os portais da transparência, criar um extrator para acessar os links atrelados às licitações, obter os dados e estruturá-los.

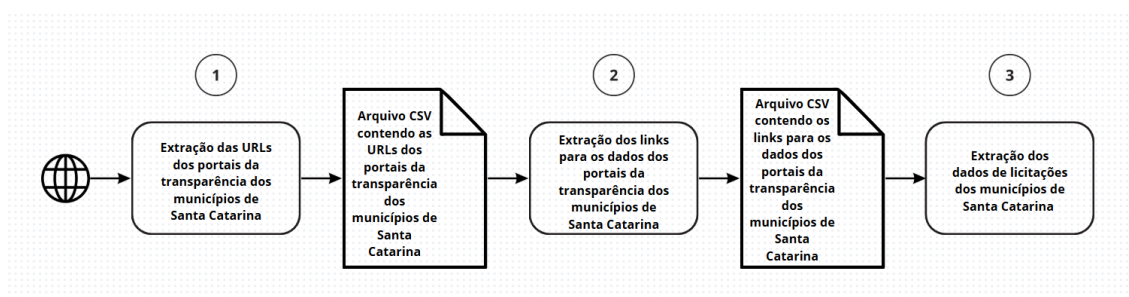


Figura 1. Representação visual das etapas do desenvolvimento.

3.1. Levantamento dos Portais da Transparência

Primeiramente, foi necessário identificar os portais da transparência das prefeituras dos municípios de Santa Catarina. Para isso, foi elaborado um processo de coleta que envolve duas etapas principais. Na primeira, foi realizada a obtenção da lista de municípios do estado a partir do site da Wikipedia. Em seguida, para cada município, uma pesquisa automatizada no Google foi executada, combinando o nome da cidade com a palavra-chave “prefeitura”. Posteriormente, foi definida uma condicional para filtrar os resultados obtidos com a pesquisa, comparando as URLs retornadas com as expressões regulares `https://(.*)\.sc\.gov\.br/` e `https://(.*)\.atende\.net/`. A primeira expressão regular possui o formato do “gov.br”, o que é um padrão em sites governamentais; porém, foi verificado manualmente a presença de portais da transparência no formato “atende.net”. Além disso, foram descartadas as URLs com “diariomunicipal” ou “img” em suas composições, evitando o site do diário municipal do estado de Santa Catarina e imagens durante a pesquisa.

Como resultado, foram obtidas as URLs dos portais de 293 municípios. Os municípios de São Lourenço do Oeste e Maravilha tiveram de ser acrescentados manualmente, pois suas URLs não foram conseguidas pela aplicação. Em posse das URLs de cada portal, foi realizada uma busca automatizada pelos portais da transparência dentro de cada URL obtida na Etapa anterior. Para isso, foram extraídos os links que continham a palavra “transparencia” em sua composição. Devido à falta de padronização no código

HTML dos portais, muitos links apresentavam a palavra “transparência” dentro de um mesmo portal; assim, múltiplos links foram coletados inicialmente para cada município.

Para organizar os resultados, foi gerado um arquivo contendo a relação entre cada município e os possíveis links para seu portal da transparência. Dessa forma, a extração resultou em um mapeamento abrangente dos possíveis portais de transparência para cada município. Mediante os links dos possíveis portais da transparência adquiridos através da implementação discutida, foi necessário averiguar os tipos de portais comuns entre os municípios para reduzir a lista a um portal da transparência por cidade - tal processo foi realizado manualmente. Conforme apresentado na Figura 4, foi possível verificar a ocorrência de 8 diferentes tipos de portal dentre todos os municípios.

O próximo passo consistiu na geração de um arquivo contendo informações sobre os portais da transparência dos municípios de Santa Catarina. Esse arquivo apresenta três colunas principais: “Portal da transparência”, que fornece o link para o portal; “Status”, que indica a situação atual da página vinculada ao link; e “Município”, que identifica o município correspondente. A Figura 2 exemplifica parte do arquivo gerado neste passo.

Portal da transparência	Status	Município
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	São Lourenço do Oeste
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	Maravilha
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	Angelina
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	Abdon Batista
https://transparencia.e-publica.net/epublica-portal/#/abelardo_luz/portal	200	Abelardo Luz
https://agrolandia.atende.net/transparencia	200	Agrolândia
https://transparencia.e-publica.net/epublica-portal/#/agronomica/portal	200	Agronômica
https://transparencia.betha.cloud/#/OnQGjBomnfW13rvf-FSXpw==	200	Água Doce
https://transparencia.betha.cloud/#/WKtxSeGNKgwqAGmqU-Y1iA==	200	Águas de Chapecó
https://aguasfrias.gov.br.cloud/pronimtb/	200	Águas Frias
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	Águas Mornas
https://e-gov.betha.com.br/transparencia/01037-157/main.faces	200	Alfredo Wagner
https://altobelavista.atende.net/transparencia	200	Alto Bela Vista

Figura 2. Exibição de parte do conteúdo do arquivo contendo as URLs dos portais da transparência.

Com o arquivo gerado, foi possível avançar para a Etapa de extração do conteúdo disponibilizado pelos portais da transparência. Considerando a existência de oito tipos distintos de sites, desenvolveu-se um algoritmo extrator específico para cada um, dado que a estrutura HTML de um determinado tipo de portal é, na maioria dos casos, idêntica ou muito semelhante.

Conforme apresentado na Figura 1, esta Etapa gerou como resultado um arquivo contendo os dados oferecidos por cada município em seu portal da transparência. Este arquivo contém as colunas “Município”, “Dado” e “Link”, possibilitando a identificação necessária para posterior extração dos dados (Etapa 3 da Figura 1), concordante ao apresentado na Figura 3.

3.2. Algoritmo de Extração

O algoritmo de extração utiliza as URLs obtidas na Etapa de coleta para acessar as páginas de licitações nos portais da transparência. Para cada tipo de dado a ser extraído (e.g., informações gerais, contratos, lances, etc.) em um determinado ano.

Município	Dado	Link
São Lourenço do Oeste	Compras diretas referente a Covid-19	https://e-gov.betha.com.br/transparencia/01037-157/con_comprasdiretasacovid.faces
São Lourenço do Oeste	Contratos referente a Covid-19	https://e-gov.betha.com.br/transparencia/01037-157/con_contratosacovid.faces
São Lourenço do Oeste	Empenhos referente a Covid-19	https://e-gov.betha.com.br/transparencia/01037-157/con_despesasempenhadasacovid.faces
São Lourenço do Oeste	Licitações referente a Covid-19	https://e-gov.betha.com.br/transparencia/01037-157/con_licitacoesacovid.faces
São Lourenço do Oeste	Receitas referente a Covid-19	https://e-gov.betha.com.br/transparencia/01037-157/con_receitasacovid.faces
São Lourenço do Oeste	Receitas	https://e-gov.betha.com.br/transparencia/01037-157/con_comparativoreceita.faces
São Lourenço do Oeste	Ingressos de receitas	https://e-gov.betha.com.br/transparencia/01037-157/con_ingressosreceitas.faces
São Lourenço do Oeste	Despesas por credor	https://e-gov.betha.com.br/transparencia/01037-157/con_gastoporfavorecido.faces
São Lourenço do Oeste	Execução de despesa	https://e-gov.betha.com.br/transparencia/01037-157/con_execucaodespesa.faces
São Lourenço do Oeste	Execução de programas	https://e-gov.betha.com.br/transparencia/01037-157/con_execucaoprogramas.faces
São Lourenço do Oeste	Gastos diretos de governo	https://e-gov.betha.com.br/transparencia/01037-157/con_comparativodespesa.faces
São Lourenço do Oeste	Gastos diretos por despesa	https://e-gov.betha.com.br/transparencia/01037-157/con_gastodiretopordespesa.faces
São Lourenço do Oeste	Gastos diretos por favorecido	https://e-gov.betha.com.br/transparencia/01037-157/con_gastodiretoporfavorecido.faces

Figura 3. Exibição de parte do conteúdo do arquivo contendo os links para os dados dos portais da transparência.

Como os portais possuem diferentes formatos e padrões, é necessário o desenvolvimento de scripts específicos para cada um, adaptando a extração de acordo com os atributos e localizações dos elementos (XPaths, seletores, etc.). Após a obtenção dos dados, estes são armazenados em arquivos estruturados para garantir a persistência. Os documentos extraídos são organizados em diretórios nomeados de acordo com a cidade e o ano de referência. Para minimizar perdas decorrentes de falhas na execução, a extração ocorre de forma segmentada, gerando arquivos estruturados ao término da obtenção dos dados de cada município. O Algoritmo 1 mostra, em alto nível, o algoritmo de extração utilizado. A Figura 4 exemplifica o processo de extração como um todo, levando em conta uma posterior integração dos dados conseguidos.

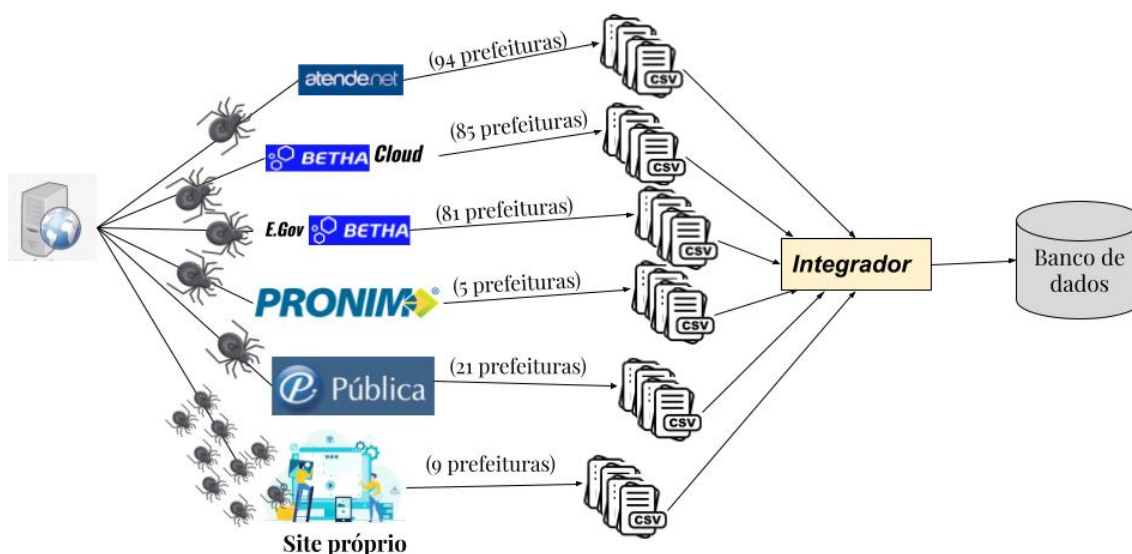


Figura 4. Representação visual da Etapa de extração.

4. Desenvolvimento

O arquivo gerado na Etapa 2 possibilitou a criação de extratores de dados referentes às licitações. O tipo de dado associado ao link possibilitou acessar a informação a ser extraída de forma mais fácil, evitando técnicas de crawling.

Algoritmo 1: Extraction of Public Procurement Data from Municipal Transparency Portals

Data: List of municipalities, data type, year, CSV file with municipalities and procurement page links

Result: CSV file for each municipality and year with extracted data

$rowsList \leftarrow LoadCSV(csvFileWithLinks)$

foreach $row \in rowsList$ **do**

$municipality \leftarrow row(municipality)$

$link \leftarrow row(link)$

if $municipality \in municipalities$ **then**

$data \leftarrow ExtractData(link, dataType, year)$

$outputFileName \leftarrow GenerateFileName(municipality, year, dataType)$

$SaveDataToCSV(data, outputFileName)$

return

Procedure $ExtractData(link, dataType, year)$:

 Access the procurement page using the provided link;

 Filter data by year and data type;

 Extract and format the relevant data;

return formatted data;

Procedure $GenerateFileName(municipality, year, dataType)$:

return $municipality + "_" + dataType + "_" + year + ".csv"$;

Procedure $SaveDataToCSV(data, fileName)$:

 Open a new CSV file with the name $fileName$;

 Write $data$ to the CSV file;

 Close the file;

Seguindo a metodologia para algoritmo de extração, primeiramente, foi definido o serviço `atende.net` para iniciar o desenvolvimento dos extratores, tendo em vista o fato de ser o mais comum entre os portais. Foi observado que cada dado oferecido por determinada licitação (e.g. contratos, vencedores, empenhos, etc.) estava disponível numa página diferente do portal, o que dificultou o processo de extração, demandando grande utilização do Selenium. Ademais, um problema recorrente com a execução da aplicação em portais `atende.net` foi a constante verificação do reCAPTCHA [Von Ahn et al. 2008]. Tal questão prejudicou a eficiência do programa, visto que provocava a necessidade de intervenções manuais para resolver os quebra-cabeças e permitir a continuidade do funcionamento.

Posteriormente, foram desenvolvidos os extratores para os sites “`pronimtb`” e portais da transparência de Blumenau e Gaspar, que não apresentaram verificações contra automatização de acesso.

4.1. Tecnologias Utilizadas

As aplicações apresentadas neste trabalho foram desenvolvidas em linguagem Python, utilizando as ferramentas Beautiful Soup, para o parsing de conteúdo HTML, e Selenium, a fim de possibilitar a interação com sites dinâmicos.

4.2. Superação de Barreira Técnicas

Em portais atende.net, a fim de lidar com a numerosa quantidade de janelas presentes nesses sites, diversos módulos tiveram de ser criados no script para lidar com o processo de chegar até o dado requerido e extraí-lo. Além disso, as verificações do serviço reCAPTCHA fizeram-se presentes na maioria das vezes em que a aplicação foi executada, inviabilizando a autonomia completa do programa. Assim, foi necessário adicionar uma espera ocupada todas as vezes que “Não sou um robô” fazia-se presente na tela, permitindo que a aplicação parasse as requisições e esperasse o usuário resolver o teste, possibilitando, assim, a continuação da extração.

4.3. Resultados Parciais

A Tabela 1 apresenta os tamanhos dos dados conseguidos até o momento, mostrando também a cidade e o(s) ano(s) de que foram extraídos. Pode-se perceber um volume maior de conteúdo nas cidades de Blumenau e Gaspar, haja vista a ausência de verificadores como reCAPTCHA, o que permitiu extrações sem intervenção manual.

Deve ser destacado que para cada ano informado na Tabela 1, arquivos CSVs de diferentes dados referentes às licitações (e.g. contratos, vencedores, empenhos, etc.) foram gerados para cada cidade, conforme previsto pela Etapa de extração.

Tabela 1. Tamanho dos dados extraídos.

Serviço de Portal da Transparência	Municípios com Dados de Licitações Extraídos	Ano(s)	Tamanho dos Dados Extraídos (MB)
atende.net	Anita Garibaldi	2021	451,8
	Arabutã		63,2
	Araquari		241,6
	Arvoredo		19,3
	Aurora		100,8
	Caxambu do Sul		273,0
	Chapadão do Lageado		167,7
	Garuva		472,7
Portal da transparência de Blumenau	Blumenau	2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2023, 2024	26624
Portal da transparência de Gaspar	Gaspar	2022, 2023, 2024	5939,2
pronimtb	Caibi	2025	87,7
	Riqueza		226,1
	Saudades		16,9

5. Discussão

Um dos principais benefícios da metodologia desenvolvida foi a centralização das informações, permitindo a consolidação dos dados de licitações em um único arquivo estruturado. Com essa abordagem, torna-se possível acessar e comparar dados de diferentes municípios sem a necessidade de consultas individuais a cada portal, proporcionando um panorama mais abrangente dos gastos públicos no estado. A Figura 5 exemplifica a unificação de dados de licitações de diferentes municípios.

Município	Licitacao	Tipo_concorrancia	Situacao	Regime_execucao	Tipo_julgamento	Tipo_comparacao	Tipo_objeto
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 4_2021	Registro de Preços	Homologada	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 3_2021	Registro de Preços	Frassada	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 2_2021	Registro de Preços	Homologada	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 1_2021	Registro de Preços	Deserta	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Inexigibilidade - Lei 8.666/1993 / 17_2021	Credenciamento/Chamamento	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Dispensa de Licitação - Lei 8.666/1993 / 16_2021	Normal	Homologada	None	None	Por Item	Obras e Serviços de Engenharia
Araquari	Inexigibilidade - Lei 8.666/1993 / 15_2021	Credenciamento/Chamamento	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 14_2021	Registro de Preços	Homologada	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Inexigibilidade - Lei 8.666/1993 / 13_2021	Normal	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Convite - Lei 8.666/1993 / 12_2021	None	Homologada	None	Menor Preco	Preço Global	Contratação de Serviços
Araquari	Dispensa de Licitação - Lei 8.666/1993 / 11_2021	Normal	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Convite - Lei 8.666/1993 / 10_2021	None	Frassada	None	Menor Preco	Preço Global	Contratação de Serviços
Araquari	Inexigibilidade - Lei 8.666/1993 / 9_2021	Credenciamento/Chamamento	Aberta	None	None	Por Item	Contratação de Serviços
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 8_2021	Registro de Preços	Homologada	None	Menor Preco	Por Item	Contratação de Serviços
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 7_2021	Registro de Preços	Homologada	None	Menor Preco	Por Lote	Contratação de Serviços
Araquari	Inexigibilidade - Lei 8.666/1993 / 6_2021	Normal	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Dispensa de Licitação - Lei 8.666/1993 / 5_2021	Normal	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Inexigibilidade - Lei 8.666/1993 / 4_2021	Credenciamento/Chamamento	Homologada	None	None	Por Item	Contratação de Serviços
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 3_2021	Registro de Preços	Anulada Total	None	Menor Preco	Por Lote	Contratação de Serviços
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 2_2021	Registro de Preços	Homologada	None	Menor Preco	Por Item	Aquisição de Bens
Araquari	Pregão Eletrônico - Lei 8.666/1993 / 1_2021	Registro de Preços	Deserta	None	Menor Preco	Por Lote	Contratação de Serviços
Araquari	Dispensa de Licitação - Lei 8.666/1993 / 4_2021	Normal	Homologada	None	None	Por Item	Contratação de Serviços

Figura 5. Exibição de parte do conteúdo de um arquivo contendo os dados gerais das licitações.

A implementação do processo trouxe ganhos expressivos na eficiência da coleta de dados. Anteriormente, a necessidade de acessar manualmente cada portal e extrair informações tornava a análise comparativa inviável em larga escala. Com o uso de técnicas de web scraping, o tempo e o esforço necessários para obtenção desses dados foram reduzidos drasticamente. Além disso, a centralização das informações permite a realização de análises estatísticas e comparativas, viabilizando a identificação de padrões de gastos e possíveis irregularidades nos processos licitatórios municipais.

Outro benefício significativo é o fortalecimento da transparência pública. Com os dados extraídos de maneira estruturada, pesquisadores, jornalistas e órgãos de controle podem acessar informações de forma mais rápida e eficiente, contribuindo para uma maior fiscalização e controle social sobre o uso dos recursos públicos. Apesar das vantagens apresentadas, a solução automatizada enfrenta desafios que limitam sua plena implementação:

- I. Dificuldade de padronização: A diversidade de plataformas utilizadas pelos municípios catarinenses para disponibilizar informações impõe barreiras à extração uniforme dos dados. Foram identificados oito tipos diferentes de portais da transparência, cada um com sua estrutura e formato próprios, exigindo o desenvolvimento de extratores específicos para cada modelo.
- II. Impacto das atualizações frequentes nos sites: As mudanças nos portais da transparência podem modificar a estrutura das páginas e, consequentemente, inviabilizar o funcionamento dos scripts de extração. Esse problema exige monitoramento contínuo e ajustes regulares nos extratores para garantir a integridade e atualização dos dados coletados.

- III. Mecanismos de segurança: Muitos portais utilizam sistemas de verificação, como o reCAPTCHA, que dificultam a automação do processo. Em diversos casos, foi necessária intervenção manual para permitir o acesso aos dados, o que reduz a eficiência e a autonomia do método desenvolvido.

6. Conclusão e Trabalhos Futuros

Este trabalho apresentou uma abordagem automatizada para extrair e centralizar dados de licitações públicas nos municípios de Santa Catarina, facilitando o acesso, a análise e a fiscalização dos gastos públicos. A metodologia fortalece o controle social ao permitir comparações e investigações detalhadas sobre o uso dos recursos. Os resultados demonstram um impacto positivo na transparência, beneficiando cidadãos, pesquisadores e órgãos fiscalizadores. No entanto, desafios como a falta de padronização dos portais, mudanças estruturais nos sites e barreiras como o reCAPTCHA indicam a necessidade de aprimoramentos contínuos.

Como trabalhos futuros, sugere-se a ampliação da metodologia para outros estados, possibilitando uma visão ainda mais abrangente dos processos licitatórios em nível nacional. Além disso, o uso de ferramentas mais avançadas de web scraping e inteligência artificial pode otimizar a extração dos dados, reduzindo a necessidade de intervenção manual e melhorando a adaptabilidade do sistema às mudanças nos portais da transparência. Essas evoluções contribuirão para o aperfeiçoamento da solução e para o fortalecimento da transparência na gestão pública.

Referências

- de Assis, W. V. and Gomide, J. V. B. (2021). Web scraping em dados públicos: método para extração de dados dos gastos públicos dos vereadores da câmara municipal de belo horizonte. *Informação & Informação*, 26(4):319–341.
- de Mello, C. (2004). *Curso de direito administrativo*. Malheiros Editores.
- de Moraes Almeida, G. H. and Silva, F. A. (2018). Web analysis: Análise de dados de transparência municipal. Technical report, Universidade Federal de Viçosa.
- Galdino, I. M., de Lima Gallindo, E., and Moreira, M. W. (2020). Utilização de bots para obtenção automática de dados públicos usando as técnicas de web crawling e web scraping. In *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*, pages 172–179. SBC.
- Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008). re-captcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468.