



Aumentando a eficiência da fiscalização social sobre gastos públicos por meio de uma aplicação web baseada em dados abertos

AFONSO S. JACINTO (IFPB, Campus João Pessoa), KYM KANATTO G. MELO (IFPB, Campus João Pessoa), MATIAS S. RIBEIRO NETO (IFPB, Campus João Pessoa), RONEI S. OLIVEIRA (IFPB, Campus João Pessoa), DAMIRES Y. SOUZA (IFPB, Campus João Pessoa)

E-mails: afonso.serafim@academico.ifpb.edu.br, damires@ifpb.edu.br, kanatto.melo@academico.ifpb.edu.br, matias.ribeiro@academico.ifpb.edu.br, ronei.santos@academico.ifpb.edu.br.

Área de conhecimento:(Tabela CNPq): 1.03.03.04-9 Sistemas de Informação.

Palavras-Chave: aplicação web; integração de dados; dados abertos.

1 Introdução

Na última década, impulsionados pela Lei Complementar 131 de 2009 - Lei da Transparência - e pela Lei nº 12.527 de 2011 - Lei de Acesso à Informação (LAI) - inúmeros entes federativos passaram a disponibilizar informações governamentais a partir de portais da transparência e portais de dados abertos, baseados na filosofia de que esses dados devem ser acessíveis para qualquer cidadão, sem restrições de uso (OPEN KNOWLEDGE FOUNDATION, 2015). Espera-se que, com a liberação desses dados para o público em geral, a administração pública possa oferecer uma maior transparência e despertar um senso de responsabilização dos atores políticos por meio da fiscalização social, de forma que a participação dos cidadãos no processo democrático seja mais pró-ativa (ATTARD *et al.*, 2015).

No entanto, a descentralização das ações de liberação de dados faz com que os conjuntos de dados sejam produzidos em diferentes formatos, gerando silos de informações, que isolam os dados de outras fontes e impedem que se criem conexões para atender a consultas de dados mais complexas. Essa limitação restringe o potencial de reuso das informações, pois dificulta a integração dos dados a partir de diferentes fontes, já que o processo de conhecimento dos conjuntos de dados, sua limpeza e descoberta de interligações com outras bases de dados fica a cargo do usuário que os consome (PENTEADO; CARLOS; ISOTANI, 2021).

Diante desse contexto, esse trabalho apresenta o resultado de um processo de coleta, preparação e integração de dados abertos governamentais de entes federativos da Paraíba, buscando, a partir das etapas citadas, prover um conjunto de dados mais completo e rico sobre gastos públicos em licitações e contratos nos últimos anos. Além disso, este artigo mostra uma aplicação web baseada no conjunto de dados produzido que disponibiliza informações processadas que possam munir a população em geral de conhecimento suficiente para aumentar a eficiência do seu poder de fiscalização sobre os atores políticos.

2 Materiais e Métodos

A metodologia aplicada neste trabalho foi de caráter exploratório e descritivo. Na fase exploratória foi realizada uma investigação aprofundada de diversas fontes de dados, identificando quais os padrões de dados existentes, quais conjuntos de dados poderiam ser extraídos e integrados de forma a construir uma coleção única de dados mais completa. Ao final do processo exploratório, foram eleitas seis fontes de dados que seriam utilizadas para a pesquisa: Portal de Dados Abertos do SAGRES - TCE/PB (FD1), Licitações do Portal da Transparência da Prefeitura de João Pessoa (FD2), Contratos do Portal da Transparência da Prefeitura de João Pessoa (FD3), Contratos do Portal de Dados Abertos do Governo da Paraíba (FD4), Contratos do Portal da Transparência do Governo da Paraíba (FD5) e Aditivos do Portal de Dados Abertos do Governo da Paraíba (FD6). Com as fontes de dados definidas foi possível iniciar a fase descritiva da pesquisa.

A fase descritiva foi realizada em cinco fases: extração dos dados brutos, preparação, integração e carga dos dados em base única, e implementação da aplicação baseada em dados (*data-driven application*). Para a extração dos dados brutos em FD1, FD2, FD3, FD4, FD5 e FD6, que juntos correspondem a cerca de 300.000 registros em formatos heterogêneos (JSON, TXT, CSV), foi utilizada a linguagem de programação Python aliada às bibliotecas Pandas, que permite manipular grande massa de dados de forma robusta, flexível e com alto desempenho, e Selenium, que permite realizar o processamento de dados simulando um navegador web, o que foi essencial para realizar o processo de *Web Scraping* em FD6. Segundo MITCHELL (2018), *Web Scraping* é uma prática de coletar dados em páginas web



utilizando uma ampla variedade de técnicas de programação e tecnologias de forma automatizada. Como saída da fase de extração de dados obtivemos três coleções de dados brutos de licitações e contratos: Sagres - TCE/PB (DB1), Governo da Paraíba (DB2) e Prefeitura Municipal de João Pessoa (DB3), no qual o DB1 foi classificado como os dados de maior confiabilidade, por serem disponibilizados pelo órgão de controle das contas públicas do estado da Paraíba. Para a preparação dos dados, foi realizada uma limpeza e padronização dos dados em DB1, DB2 e DB3, ou seja, foram padronizados formatos de exibição (datas, chaves primárias, número de licitação, número de contrato, número de processos, entre outros), codificação (UTF-8, ISO-8859-9) e tipos de dados (*String*, *Array*, *Object*, *Double*, *Date*, *Integer*). Também foi realizada uma remoção de valores nulos e dados considerados irrelevantes para os requisitos da aplicação. Na fase de integração de dados foi utilizado o sistema NoSQL MongoDB. Nele foram executados os scripts de integração no MongoDB Shell, mongosh, que é um ambiente REPL JavaScript e Node.js totalmente funcional para interagir com implantações do MongoDB. Esses scripts relacionaram os dados brutos DB1, DB2 e DB3 de forma a encontrar os dados que se interconectam entre as três bases. Na fase de carga, os dados encontrados na fase anterior foram persistidos em uma base única de licitações e contratos (BULC) no MongoDB.

Na quinta e última fase, foi desenvolvida uma aplicação web baseada nos dados da BULC. O *Back-End* da aplicação foi desenvolvido como uma API (*Application Programming Interface*) utilizando a linguagem de programação Java com o *framework* Spring Boot. Para o desenvolvimento do *Front-End* foi utilizado o *framework* Angular e o kit de ferramentas de *Front-End* de código aberto para a construção de interfaces responsivas, Bootstrap. Para melhorar o tempo de resposta das consultas, um servidor web intermediário Redis foi utilizado entre o *Front-End* e o *Back-End*, esse servidor funciona com cache em memória armazenando chave e valor para disponibilizar como retorno da consulta dos dados. Todos os *frameworks* utilizados são comerciais e disponibilizados de forma gratuita para o desenvolvimento de sistemas web.

3 Resultados

Esta pesquisa traz como resultados duas importantes contribuições. A primeira é uma base de dados integrada que dispõe de dados de licitações e contratos de forma mais completa em um único ponto de acesso, sendo essa também disponibilizada em forma de API para que seja explorada por terceiros, como pode ser verificado na Figura 1.

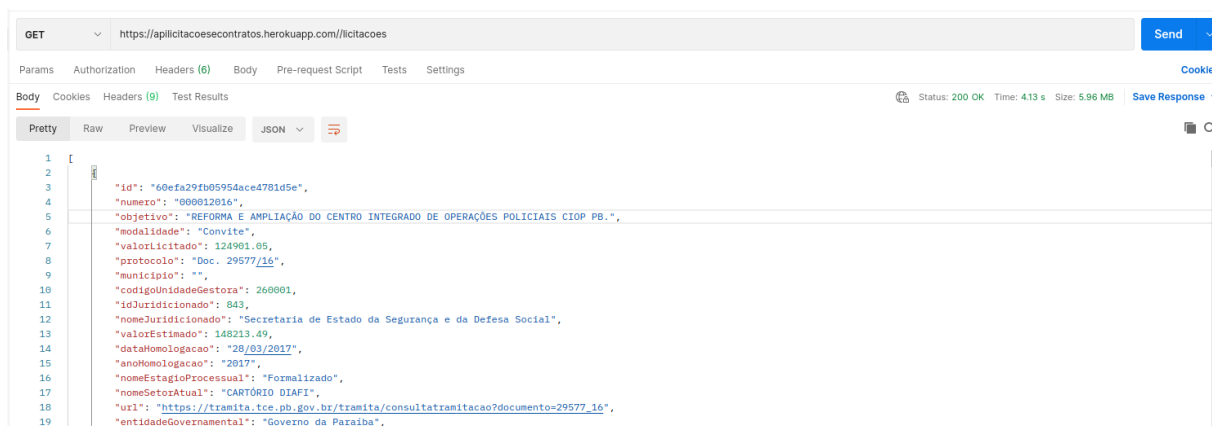


Figura 1: *Endpoint* para a API de consulta a base integrada.

A segunda importante contribuição é uma aplicação baseada nos dados integrados que permite que a sociedade civil possa consultar informações de licitações e contratos do Governo da Paraíba e Prefeitura Municipal de João Pessoa entre os anos de 2017 e 2021. O cidadão pode consultar informações simples e objetivas como os valores das licitações e contratos durante os anos, como podemos observar na Figura 2. É possível identificar também, por meio de gráficos, quais os cinco órgãos que mais realizaram licitações nos anos abordados. Também são apresentadas as unidades que mais realizaram licitações dispensadas por conta da COVID-19. Além disso, é possível buscar de forma simples e direta, por meio de filtros, os dados de fornecedores, contratos e licitações.



Figura 2: Visualização de gastos com licitações durante os anos.

O foco da aplicação é trazer informações de forma simples e objetiva e que seja de fácil consulta e interpretação para o cidadão em geral, sem que o mesmo necessite ter um prévio conhecimento tecnológico ou técnico. Espera-se, com isso, que as pessoas possam identificar, por exemplo, possíveis fraudes aplicadas com o dinheiro público, como na pesquisa de Fraga (2017), na qual foram identificados padrões de dados para grupos de empresas suspeitas de práticas anticompetitivas em licitações ocorridas nos municípios paraibanos entre 2005 e 2016. Nessa pesquisa foi utilizada a fonte de dados FD1, que também compõe a nossa aplicação web.

4. Considerações Finais

A presente ferramenta pretende se tornar uma importante contribuição do IFPB, Campus João Pessoa, para o desenvolvimento de soluções tecnologicamente inovadoras para a transformação digital na gestão pública, pois a plataforma desenvolvida permite uma comunicação visual rápida de um grande volume de informações, tornando-se uma ferramenta essencial para apoiar tomadas de decisão baseadas em dados. Por sua vez, observa-se como melhorias futuras a possibilidade da inserção de novas fontes de informação e de novas formas de visualização dos dados, assim proporcionando a geração de novos contextos de informações.

Referências

ATTARD, J.; ORLANDI, F.; SCERRI, S.; AUER, S. A systematic survey of open government data initiatives. *Government Information Quarterly*, vol. 32, no. 1, pp. 399-418, 2015. Disponível em: <http://dx.doi.org/10.1016/j.giq.2015.07.006>. Acesso em: 17 jul. 2021.

FRAGA, A. A. Detecção de casos suspeitos de fraudes em licitações realizadas nos municípios da Paraíba: uma aplicação de técnicas de mineração de dados. 2017.

MITCHELL, R. Your First Web Scraper. In: *WEB scraping with Python: Collecting more data from the modern web*. 2. ed. rev. Sebastopol: O'Reilly Media, Inc., 2018. cap. 1, p. 3-13.

OPEN KNOWLEDGE FOUNDATION. Defining open in open data, open content and open knowledge. *Open Definition*, v. 2.1, 2015. Disponível em: <https://opendefinition.org/od/2.1/pt-br/>. Acesso em: 17 de julho de 2021.

PENTEADO, B.; CARLOS, M. J.; ISOTANI, S. Process Model with Quality Control for the Production of High Quality Linked Open Government Data. *IEEE Latin America Transactions*, [S. l.], v. 19, n. 3, p. 421-429, 2021. Disponível em: <https://latam.ieee9.org/index.php/transactions/article/view/3501>. Acesso em: 17 Julho de 2021.