Novo artigo sobre a ferramenta pyHDB!

2023-01-08

{{< image src=“/img/pyhdb.png” alt=“logo pyHDB” position=“left” style=“border-radius: 8px;” >}}

No último dia de 2022 saiu novo número da Revista [História da Historiografia](https://www.historiadahistoriografia.com.br/revista/issue/view/45) com um artigo meu chamado [“pyHDB - Ferramenta Heurística para a Hemeroteca Digital Brasileira: utilizando técnicas de web scraping para a pesquisa em História”](https://doi.org/10.15848/hh.v15i40.1904).

É meu xodó!

No artigo apresento a pyHDB, ferramenta de suporte metodológico para a pesquisa na [Hemeroteca Digital Brasileira](http://memoria.bn.br/hdb/periodico.aspx), da Biblioteca Nacional. Analiso sua relação com teoria, metodologia e epistemologia da história, focando na seleção, coleta, organização e critica das fontes.

A pyHDB é um programa em linha de comando escrito em Python que gera relatórios, planilhas e data sets a partir dos resultados da Hemeroteca Digital Brasileira. Os parâmetros de Local, Período e Termo da busca são definidos pelo usuário e a pyHDB busca e inicia o processo de coleta de informações.

{{< image src=“/img/pyhdb\_par.png” alt=“parameters” position=“left” style=“border-radius: 8px;” >}}

São registrados os parâmetros, a hora e data da busca, os resultados, o nº de ocorrências, o nº de páginas de cada acervo, a frequência de ocorrências por páginas, as informações de cada ocorrência, relatórios de erros e o download da imagem quando legalmente permitido.

{{< image src=“/img/pyhdb\_scrape.png” alt=“scrape” position=“left” style=“border-radius: 8px;” >}}

Veja um exemplo dos dados coletados e organizados pela pyHDB em um arquivo .csv:

{{< image src=“/img/pyhdb\_sheet.jpeg” alt=“csv” position=“left” style=“border-radius: 8px;” >}}

Aqui a estrutura de diretórios criada pela pyHDB:

{{< image src=“/img/pyhdb\_dir.png” alt=“directories” position=“left” style=“border-radius: 8px;” >}}

E as imagens baixadas:

{{< image src=“/img/pyhdb\_imgs.png” alt=“images” position=“left” style=“border-radius: 8px;” >}}

Um dos exemplos utilizados no artigo foi Monteiro Lopes. Os dados confirmam o intenso debate em torno das eleições do advogado negro em 1903 para o conselho municipal do RJ e em 1909 como deputado federal.

{{< image src=“/img/pyhdb\_graph.jpeg” alt=“Monteiro Lopes” position=“left” style=“border-radius: 8px;” >}}

Pra fechar com chave de ouro, [o parecer incrível do Pedro Silveira sobre o artigo também foi publicado na revista](https://www.historiadahistoriografia.com.br/revista/article/view/2071). Não é apenas uma leitura do meu artigo, mas uma reflexão sofisticada sobre o próprio fazer historiográfico na era digital.

Agradeço aos membros do LabhdUfba, especialmente a Gabriel Andrade, pelas dicas e ajuda com Python, e ao seu coordenador, Leonardo F. Nascimento, pela parceria, leitura e por ter sugerido o nome pyHDB. Agradeço a Yaci Farias e Luara Santos pela leitura e comentários durante a escrita do artigo. Agradeço a Ana Carolina Veloso e Priscila Valverde Silveira pelo trabalho desenvolvido como bolsistas de iniciação científica no curso de História da Unilab, campus dos Malês ao longo de 2020 e 2021. Agradeço a Alexandra Elbakyan por sua atuação na ciência livre. Gostaria de agradecer aos membros do conselho editorial da revista pelo apoio e aos revisores anônimos pelas valiosas sugestões.

Toda a documentação está disponível na página da [pyHDB](https://ericbrasiln.github.io/pyHDB/).

*Data set* utilizados no artigo: https://doi.org/10.5281/zenodo.5699000

Gráficos interativos, imagens e scripts: https://ericbrasiln.github.io/analise\_pyHDB/

Qualquer questão entrem em contato ou abram uma issue lá no [Github](https://github.com/ericbrasiln/pyHDB/issues).