
```
caminho =  
input(“Insira o  
caminho para  
História Digital:”)
```

Conferência de abertura

Eric Brasil

17 de outubro de 2023

historia_digital.py

Quando estamos programando, é comum precisarmos passar o caminho de um arquivo para o programa que estamos criando. Se quisermos ler uma fonte, abrir uma imagem, analisar dados de uma planilha ou fazer o download de um pdf para uma pasta específica, temos que informar ao computador qual o seu caminho. Ou seja, onde o arquivo está armazenado em sua máquina.

Ele necessariamente deve ser preciso: explícito e completo, desde sua pasta raiz até o nome do arquivo com sua respectiva extensão.

Por exemplo, para ler um arquivo Python (.py) chamado historia_digital, eu passaria o seguinte comando com Python em execução:

SLIDE 1

```
1 >>> open('historia_digital.py', 'r').  
      read()
```

O retorno seria o conteúdo do arquivo `historia_digital.py` impresso em minha tela.

Porém, caso o caminho para esse arquivo não esteja correto, o comando não terá sucesso e um receberemos um aviso de erro.

Mas calma, essa conferência não vai ser um tutorial de Python, não precisam se desesperar! É apenas o bom e velho recurso à metáfora para iniciar a fala.

Pretendo, hoje fazer uma reflexão coletiva com vocês sobre aspectos, que na minha avaliação, são urgentes para nossa disciplina: a construção metodológica de pesquisas com dados e ferramentas digitais e sua relação com a construção do conhecimento histórico e o próprio caráter científico da disciplina.

Para isso, farei um breve histórico do uso da compu-

tação na pesquisa histórica, antes e depois do termo História Digital; apontarei questões gerais envolvendo as práticas de pesquisa atuais, com o exemplo da hereroteca digital brasileira e como seus aspectos técnicos e vieses de seu acervo impactam epistemologicamente na História; e por fim refletir sobre caminhos possíveis e necessários para o futuro e o presente.

Mas antes, justificando minhas escolhas para hoje, voltemos à metáfora.

SLIDE 2

**Arquivo ou pasta não existe:
'historia_digital.py'**

Ao executar o comando anterior, recebi o seguinte erro:

```
1 >>> FileNotFoundError: [Errno 2] No  
such file or directory:
```

```
2 'historia_digital.py'
```

Ou seja, o Python não foi capaz de encontrar o arquivo. Portanto, não pode executar suas tarefas, pois o caminho que eu passei não aponta de forma exata para o local onde o arquivo está armazenado em meu computador. Imaginando que o mesmo estivesse salvo na pasta **USP** que por sua vez está armazenada na pasta **Documentos**, o caminho correto desse arquivo seria: `~/Documentos/USP/historia_digital.py`. Agora o python será capaz de encontrá-lo e executar as tarefas desejadas.

SLIDE 3

```
1 >>> open('~ / Documentos / USP /  
    historia_digital.py', 'r').read()  
2 >>> 'print("Olá, História Digital!")'
```

E o que isso tem a ver conosco?

Assim como o python precisa de uma informação pre-

cisa para encontrar um arquivo e executar suas tarefas, nossas pesquisas, cada vez mais mediadas, viabilizadas e mesmo engendradas por ferramentas, métodos e dados digitais, demandam precisão e transparência metodológica mais acentuada e desta precisão, e do seu respectivo registro, e da transparência desses processos depende a própria possibilidade de execução, avaliação, e compreensão da pesquisa.

Em outras palavras, o desenvolvimento e aplicação de metodologias documentadas e precisas são fundamentais para a produção do conhecimento. Isso parece óbvio, porém o que percebemos, e nisso eu me incluo e imagino que falo pela maioria de nós, é que as pesquisas mediadas e baseadas em dados digitais têm, no melhor dos cenários, apenas emulado padrões metodológicos analógicos sobre um universo de recursos digitais. O que gera comumente resultados inverificáveis, não-reproduzíveis, e dificilmente

escrutináveis.

Uma pesquisa que se utiliza de buscas digitais em repositórios, como a HDB, não pode prescindir do registro detalhado dos parâmetros utilizados e da publicização tanto das escolhas feitas como dos resultados obtidos. Caso contrário, o conhecimento histórico produzido estará sempre subordinado aos vieses e direcionamentos de conjuntos de decisões técnicas e políticas que formam as páginas de buscas que utilizamos.

Como numa linguagem de programação, sem o registro preciso do caminho que leva desde a raiz até o arquivo final, nossa pesquisa em História está sujeita ao constante **erro** do caminho não encontrado.

SLIDE 4

Quais caminhos até chegar aqui?

Há um emaranhado de caminhos trilhados e outros tantos abandonados para eu chegar aqui, sentado nessa mesa tão importante, usando um blazer pra fingir que eu sou sério. Caminhos da própria relação entre história e o digital, entre nosso fazer historiográfico e a computação, mas também dos meus caminhos pessoais.

Mais do que uma *ego-história* (vou poupar vocês disso) e tentando escapar da ilusão biográfica, alertada por Bourdieu, vou falar brevemente dessa trajetória individual, pois possibilita refletir sobre mudanças que de alguma forma afetam a todos e todas.

Meus caminhos pessoais: do analógico ao digital à programação

A minha formação foi toda em História, na Universidade Federal Fluminense, lá em Niterói, no Rio de Janeiro. Os meus estudos iniciais foram bastante ligados à história social da cultura. Eu trabalho, sobretudo, com pesquisas que buscam compreender as experiências sociais das pessoas que faziam carnavais no fim do século 19 e no início do século 20, inicialmente no Rio de Janeiro, mas as pesquisas avançaram pelas afro-américas. E desenvolvi pesquisas com enfoque transnacional, analisando as experiências de populações negras no Rio de Janeiro e Trinidad e Tobago no Caribe. A minha pesquisa é muito ancorada numa história social thompsoniana buscando olhar para cultura e as relações materiais por trás daquilo que possibilita a atuação pública dessas pessoas. Trabalho

primordialmente com jornais e muita documentação policial.

O doutorado foi um momento de maior reflexão metodológica, a partir de um conjunto de dados digitais e ferramentas digitais que eu comecei a utilizar de forma intuitiva e pouco criteriosa - e eu acho que isso é uma coisa comum para a grande maioria dos pesquisadores contemporâneos, que estão hoje na ativa: digitalizar, transcrever, criar bancos de dados, tentar criar formas de analisar e conectar os dados. Isso se soma ao fato de ter conhecido a HDB.

Após a defesa do doutorado, em 2016, eu me tornei professor da UNILAB - Universidade da Integração Internacional da Lusofonia Afro-brasileira, no Campus dos Malês, na Bahia onde sou professor de História da América até hoje. E aí eu acabei me aprofundando nesse debate, sobretudo com o professor Leonardo Nascimento, que era professor de metodologia da UNI-

LAB àquela altura, e que depois passou para UFBA, onde montou o Laboratório de Humanidades Digitais do qual eu também sou membro. E eu não tenho dúvida que foi ali com os debates, diretamente com ele, que é especialista já há muito tempo em Sociologia Digital que eu percebi que muitas daquelas questões que eu estava levantando não eram novas, necessariamente, e já havia uma literatura importante por trás, mas que a gente, partindo das nossas experiências, do Brasil, das pesquisas de história social, tinha muito a contribuir. Então a partir daí que eu entrei de vez em 2017-18 na pesquisa de História Digital, entendendo o que eram as Humanidades Digitais. E estou aí, há uns cinco anos, pesquisando isso. Tanto que quase não me chamam mais para falar de carnaval. Agora só me chamam para falar de história digitais. Eu gosto também, mas de vez em quando eu fico com saudade dos meus carnavalescos.

SLIDE 5

Breve história do uso da computação na História: de Busa ao Programming historian

Eu acho que todo mundo que inicia, ou se depara pela primeira vez com um debate sobre Humanidades/História Digitais, tende a perceber-lo como muito recente, sobretudo por estarmos mergulhados na era da digitalização da relações sociais. Eu também fui mordido por esse bichinho da novidade. Porém, sendo historiador, a gente faz logo um levantamento bibliográfico pra ver o que já se falou sobre o tema. E então descobrimos que há uma história mais longa e complexa da relação entre computação e história.

Não pretendo fazer uma genealogia do termo ou falar de todas as vertentes, apenas destacar a necessidade de maior atenção para essas histórias.

Se o termo Humanidades Digitais é recente para a história, porque ele vai surgir já no século 21 e vai se popularizar muito na década de 2010, existe uma longa história do uso dos computadores nas pesquisas em humanidades, e isso remonta lá aos anos 1950, com famoso e sempre citado projeto do padre Busa, em parceria com a IBM: uma codificação da obra de São Tomás de Aquino. Isso começa com um trabalho gigante de criação de cartões perfurados. Hoje esse é um trabalho reconhecido como pioneiro e tinha um grande aporte tecnológico da IBM.

Além disso, tem muitos outros caminhos de utilização dos computadores na História e nas Humanidades. Nos anos 1960, tivemos o avanço das pesquisas da chamada Cliometria e nas décadas seguintes, sobretudo nos anos 1980, surgem várias revistas específicas tratando do uso de computadores para a pesquisa histórica, muitas delas ligadas a História Medieval, a

criação da Association for History Computing, o desenvolvimento de pesquisas utilizando bancos de dados, etc.

A partir dos anos 1990, com o boom das digitalizações em massa, o avanço da internet e a popularização dos computadores pessoais e dos programas com interface gráfica (proprietários), há um crescimento de pesquisas linguísticas utilizando recursos computacionais, buscando a compreensão de grandes conjuntos de dados digitalizados a partir de livros, por exemplo. Com a virada para o século XXI, com a web 2.0, a ampliação das interfaces gráficas para usuários e depois o surgimento das redes sociais, há um fortalecimento do campo das humanidades digitais, ao mesmo tempo em que há uma diminuição da participação de historiadores e historiadoras na área.

Agora, obviamente esse não é um movimento linear e harmonioso. São campos em intensa disputa e uma

diversidade tão grande que muitos dizem que o termo humanidades digitais não serve para nada. Muitos caminhos foram esquecidos, silenciados e outros consolidados em função de poder simbólico e capital.

É por isso que concordo com Adam Crymble, em seu livro *Technology & the Historian*, que fala que a gente precisa, mais do que definir a história digital, (e eu acho que isso serve para as Humanidades como um todo), a gente precisa fazer as suas histórias (p. 18). Seria importante fazer essas histórias para o Brasil também.

Mas o que eu queria reforçar aqui é o fato de que a consolidação do campo das Humanidades Digitais hoje está profundamente atrelada às desigualdades globais e explorações planetárias do capitalismo dominante. É uma área ainda profundamente anglófona, dependente de investimentos elevados em estrutura e formação. Caso acreditemos que só é possível desen-

volver pesquisas de ponta com esse modelo, estaremos reforçando uma visão eurocêntrica da disciplina, o que estamos a décadas buscando romper.

Um movimento importante é a busca pelo desenvolvimento de pesquisas multilinguísticas nas humanidades digitais, como o Programming Historian, que atualmente publica sua revista em inglês, espanhol, francês e portugueses, buscando publicar artigos metodológicos produzidos a partir de problemas de pesquisa reais.

SLIDE 6

História Digital existe?

Mas dito tudo isso, o que caracteriza uma história digital (toda história não é digital hoje? podemos discutir isso depois)?

Uma volta aos números? Um primeiro ponto que gostaria de chamar atenção é o risco de atrelarmos a pesquisa histórica a análise distante de grandes bases de dados, aplicando métodos e ferramentas digitais capazes de gerar visualizações belas, estatísticas complexas, mas se afastando da análise das relações sociais humanas, da busca pela compreensão das escolhas e decisões dos sujeitos históricos. Em outras palavras abdicando de uma história vista de baixo para uma história vista pelos números.

Uma crítica muito interessante a partir dessa perspectiva pode ser ouvida no podcast História Pirata #109 - História Digital Vista de baixo, com a participação do Tiago Gil da UNB (<https://open.spotify.com/episode/0JibQkhfTvhndEgTvkVLh4?>

Entendo que precisamos sofisticar nossas reflexões teóricas para utilizar de maneira condizente e eficaz as técnicas e métodos digitais. O que me leva ao segundo

ponto:

Prática e Teoria? Sob o guarda-chuva da História Digital estão abrigadas pesquisas com perspectivas teóricas muitíssimo variadas. O que as une? apenas o fato de usar alguma ferramenta digital? u

Daniel Alves (NOVA) define história digital como uma comunidade de práticas. A necessidade por utilizar, desenvolver, testar e compreender métodos e ferramentas digitais tem agrupado pesquisadores de diferentes correntes em laboratórios, centros, oficinas, treinamentos, etc.

Esse movimento tem acontecido no Brasil, acelerado pela pandemia sem dúvida, e temos visto surgir laboratórios, grupos de pesquisa, publicações, seminários etc em diversas regiões do país.

Gostaria de combinar a perspectiva da “comunidade

de práticas” com a *Hermenêutica Digital*, como definida por Fickers e Clavert:

A combinação de competências digitais críticas com uma abordagem auto-reflexiva é chamada de hermenêutica digital: explicitar como a produção do conhecimento histórico por meio de ferramentas e tecnologias digitais é o resultado de um processo complexo de interação homem-máquina, de co-construção do “objeto epistêmico” de inquirição e investigação histórica (Fickers e Clavert, 2021).

Ou seja, fazer história digital demanda não apenas utilizar tecnologias digitais na pesquisa, mas assumir deliberadamente uma postura crítica, reflexiva e explícita na correlação profunda entre técnica, método, teoria e epistemologia da história.

SLIDE 7

A técnica afetando a epistemologia

Consequentemente, o uso ingênuo ou pouco crítico de ferramentas digitais impacta diretamente na construção do conhecimento científico - parece óbvio para vocês?

A mediação das plataformas e motores de busca

Mas a pergunta que eu queria fazer é: Sem conhecimento técnico não teremos como fazer nosso trabalho?

Essa é uma questão que merece uma conversa mais atenta. Para exercermos nossa capacidade criativa e crítica enquanto historiadores/as, precisamos conhecer minimamente as estruturas, organizações, funcionamento dos arquivos que trabalhamos; precisamos conhecer minimamente as capacidades técnicas dos

gravadores e câmeras digitais que utilizamos para realizar entrevistas (ou do celular utilizado); etc.

Ao lidarmos com ferramentas ou acervos digitais ou digitalizados também precisamos conhecer a base técnica de sua formação para podermos explorá-los de forma eficiente, e mais, alargar as possibilidades epistemológicas de seu uso.

Ou seja, ao abrir um navegador e pesquisar por jornais digitalizados, precisamos entender minimamente como a lista de resultados será apresentada, e mais, como ela vai diferir enormemente para cada conjunto de variáveis combinadas: qual seu histórico, qual buscador, quais parâmetros utilizados, seu local de busca, etc.

Mas se isso já é impactante numa simples busca genérica em buscador, imagina em um repositório digital que irá basear, e muitas vezes direcionar, sua pesquisa acadêmica!

SLIDE 8

Um caso emblemático: a HDB, seus vieses e potencialidades

Vejamos o caso da Hemeroteca Digital Brasileira da FBN. Nos últimos anos tenho produzido reflexões teóricas a partir da HDB e (ontem) encerrei meu estágio de pós-doutorado justamente sobre hemerotecas de jornais digitalizados em português, realizando uma análise comparativa entre os casos brasileiro e português e produzindo um framework específico para pesquisa-as que lidam com fontes desse tipo.

Além deste motivo, recentemente tenho recebido muito pedidos de parecer e visto muitos artigos de História da Educação que utilizam a HDB como fonte, e muitas vezes como fonte principal.

Por esses motivos, entendo que ela serve como um

exemplo muito rico para pensarmos sobre os impactos de aspectos técnicos e dos vieses dos acervos para a pesquisa histórica.

SLIDE 9

O que é a HDB?

O que é?

Como funciona?

Cadê as citações?

SLIDE 10

HEMDIG(pt)

Pesquisa de pós-doc.

Framework completo para trabalhar com jornais digitalizados em português.

SLIDE 11

Visão geral do acervo Quando pesquisamos nela, temos noção precisa do escopo desse acervo? Após solicitar os dados diretamente à FBN, obtive acesso à um arquivo XML, exportado internamente da plataforma da FBN, contendo todos os dados sobre periódicos digitalizados disponíveis na HDB. Esse XML é muito rico e fundamental para a compreensão do carácter e vieses desse acervo tão importante para a pesquisa histórica brasileira. Após consulta aos setores responsáveis da FBN, recebi autorização para tornar público esses dados. Em minha pesquisa de pós-doutorado, realizei a análise desses dados e apresento para vocês alguns aspectos para nossa reflexão.

SLIDE 12

Volume de dados e variedade;

Coluna	Quantidade
Título	7685
Subtítulo	3666
Local	7608
Período	7265
Editora	5675
Periodicidade	5901
Idioma	6614

SLIDE 13

Escopo cronológico e geográfico;

- Valores únicos na coluna Local: 673

- 7685 títulos
- Cronologia: de 1521 a 2022

SLIDE 16

Viés temporal?

- Há uma concentração de periódicos digitalizados em algum período específico?

Scatter plot com datas de início: `file:///home/ebn/Documentos/framework/repositorios/BND_BR/charts/datas_inicio_scatter.h`

SLIDE 17

Gráfico de barras de datas de início e fim por décadas:

`file:///home/ebn/Documentos/Github/hemdig-framework/repositorios/BND_BR/charts/decadas_inicio_termi`

SLIDE 18

Viés geográfico?

- A HDB representa um acervo geograficamente centralizado?

Gráfico de barras com as 20 cidades mais recorrentes:

file:///home/ebn/Documentos/Github/hemdig-framework/repositorios/BND_BR/charts/cidades_bar.html

SLIDE 19 Pizza cidades com mais de 1%: file:///home/ebn/Documentos/Github/hemdig-framework/repositorios/BND_BR/charts/cidades_pie.html

SLIDE 20

- O que isso quer dizer?

A cidade do Rio de Janeiro corresponde a 28,43% do total e 47,1% entre as cidades com pelo menos 1%. A cidade de São Paulo vem em seguida, mas com apenas 3,82% do total e 6,6% entre as cidades com pelo menos 1%. A diferença entre São Paulo e as seguintes é pequena: Florianópolis com 3,1%; Recife com 2,69%; Salvador com 2,49% e Maceió com 2,40%.

Apesar da centralidade da cidade do Rio de Janeiro

nos números absolutos do acervo, é interessante notar que o estado do Rio de Janeiro possui apenas 48 cidades entre as localidades dos periódicos, atrás de Santa Catarina (49), São Paulo (59) e Minas Gerais, que apresenta o maior número de cidades no dataframe, com 190.

Cidades por Estado: file:///home/ebn/Documentos/Github/herframework/repositorios/BND_BR/charts/estados_bar.html

De maneira geral, percebemos que o acervo apresenta uma centralidade para periódicos publicados na cidade do Rio de Janeiro, entre os anos 1860 e 1920; periódicos majoritariamente publicados em língua portuguesa; com periodicidade variada, mas com destaque para periódicos semanais e mensais. Também podemos perceber que a vida da maioria dos periódicos foi bastante efêmera, com 5694 periódicos tendo apenas um ano de publicação registrado no acervo de um total de 7685. Isso corresponde a 74,1% dos periódicos.

Apenas 1991 periódicos possuem mais de um ano de publicação registrado no acervo, o que corresponde a 25,9% do total.

Paremos de usar a HDB?!

Nunca!

Entretanto, nossa formação enquanto pesquisadores precisa se transformar em aspectos técnicos e teóricos que acompanhem as transformações nas formas de acesso aos vestígios do passado.

Mas também os repositórios/arquivos precisam tornar suas escolhas técnicas (sempre políticas) e as características dos seus acervos públicas e escrutináveis. Para isso seria fundamental políticas públicas de investimento em estrutura, formação e valorização dos profissionais desses acervos.

Futuro e presente

Formação técnica

Reforço metodológico + interdisciplinaridade + laboratórios

Reflexões epistemológicas

Teoria da história mais crucial do que nunca.

Colaboração, transparência, ciência aberta

the resistance to citing digital sources that is still widespread in the humanities. We have found that this resistance is particularly prevalent where a print version of the source also exists (Blaney, 2017, parágrafo 5)

Financiamento, políticas públicas

Desigualdade global do capitalismo + desigualdades raciais e de gênero (um campo ainda marcado por privilégios)