

## **Métodos, ferramentas e repositórios digitais em português: um *framework* para a pesquisa em periódicos digitalizados**

**Docente:** Professor Doutor Eric Brasil Nepomuceno

---

### **DADOS DO CURSO**

**Curso:** História

**Instituição:** Universidade Nova de Lisboa (Lisboa, Portugal)

**Período:** de 17/10/2022 a 16/09/2023.

**Supervisor:** Professor Doutor Daniel Alves

**Descrição das competências a serem desenvolvidas durante o afastamento, alinhadas à área de atuação do servidor na Unilab, em atendimento ao disposto no §3º, do Art. 22 do Decreto Nº 9.991/19:**

Desenvolvimento de atividades de pesquisa, envolvendo a produção de conhecimento científico original, o estabelecimento de parcerias profissionais capazes de favorecer a internacionalização da UNILAB, além do desenvolvimento de ferramentas e métodos digitais para pesquisa no campo das Humanidades Digitais em português, que estejam alinhados com as propostas epistemológicas que regem o projeto institucional da UNILAB.

---

### **Introdução**

Experimentamos, ainda que de forma desigual, uma profunda virada digital nas humanidades nos últimos vinte anos (BRESCIANO, 2000, p. 7–9) e, também, modificações na História enquanto disciplina. Para a pesquisa histórica esta mudança reside na combinação entre a digitalização das fontes primárias, a profusão na criação e disponibilização de novas fontes nativamente digitais, a *dataficação* das relações sociais (SOUTHERTON, 2020) e, consequentemente, a transformação nos métodos de pesquisa e construção do conhecimento histórico.

Apesar de já possuímos uma literatura robusta sobre o uso de tecnologias digitais e computação no campo da História<sup>1</sup>, percebemos que o interesse e urgência em torno de tais questões foram aceleradas

---

<sup>1</sup>ROMEIN, C. Annemieke; KEMMAN, Max; BIRKHOLZ, Julie M.; et al. State of the Field: Digital History. *History*, v. 105, n. 365, p. 291–312, 2020.

e aprofundadas, mesmo que de forma açodada e desequilibrada, pela pandemia de COVID-19 que assola a humanidade desde o início de 2020.

Por um lado, este debate tem muito a crescer a partir do momento em que for encarado por historiadores e historiadoras das mais distintas filiações teóricas, buscando colaborar ativamente na construção de reflexões junto a arquivistas, bibliotecários, programadores, cientistas de dados. Em outras palavras, atuar no que Max Kemman chama de “zonas de troca” (KEMMAN, 2021) buscando ampliar tanto os aspectos técnicos quanto epistemológicos do desenvolvimento e uso de ferramentas e métodos digitais para a pesquisa em História.

Por outro, é fundamental que tais debates sejam enfrentados e ampliados pela produção intelectual dos países da América Latina e da África (REF), que o escrutínio sejam feito também por uma variedade ampla de idiomas e epistemologias para além da produção intelectual hegemônica, que é majoritariamente anglófona.

O cenário atual de desigualdades no acesso a tecnologias, infraestrutura de pesquisa, investimentos estatais e privados, e conhecimentos técnicos, ou ainda a predominância de ferramentas e programas em inglês e de código fechado, proprietárias, coloca desafio ainda maior para pesquisadores e pesquisadoras do Sul Global: os procedimentos básicos que devemos tomar com as fontes históricas, desde sua busca, seleção e coleta, até o tratamento e análise, tem esbarrado constantemente em tecnologias digitais. Porém, seu uso demanda recursos, conhecimentos e infraestrutura que comumente são de difícil acesso.

O ensino “de competências digitais nas várias disciplinas das Humanidades no ensino universitário” (ALVES, 2021, p.5) é central para o aprofundamento teórico e metodológico da inter-relação entre o que é próprio da História e o que é próprio da computação. Como afirma Daniel Alves (2021), o ensino das Humanidades Digitais avançou no mundo anglófono ao longo da década de 2010. Contudo, ao analisar o caso da pesquisa e ensino da História nas universidades em Portugal, entende que “a prática tem-se sobreposto à teoria, a investigação ao ensino e o trabalho em comunidade à institucionalização acadêmica” (p.7).

Nesse projeto, enfrento o problema comum a todos/as que executam a operação historiográfica: o processo de selecionar, recolher, organizar as fontes primárias e realizar sua crítica – a chamada heurística das fontes. Agora, com um aspecto diferente, uma heurística em ambientes digitais, através de ferramentas e dados digitais. O foco aqui está na análise de repositórios e interfaces gráficas que permitem o acesso a periódicos em língua portuguesa digitalizados.

Pretendo analisar as interfaces gráficas e possibilidades de acesso aos dados e sua relação com teoria, metodologia e epistemologia da História. A ideia é aproximar os saberes fundamentais da pesquisa em história com conhecimentos técnicos de programação, atuando em zonas de troca, interdisciplinares e colaborativas. Ou ainda, parafraseando Fickers e Clavert, busco “combinar habilidades digitais críticas com uma abordagem auto-reflexiva”, prática chamada de hermenêutica digital, ou seja:

Tornar explícito como a produção de conhecimento histórico através de ferramentas e tecnologias digitais é o resultado de um processo complexo de interação humano-máquina, de co-construção do 'objeto epistêmico' da inquirição e investigação histórica. (FICKERS, CLAVERT, 2021, parágrafo 6, tradução minha)

Entendo que as novas formas de realizar pesquisas em repositórios e interfaces digitais de buscas impactam, mediam e direcionam tanto a coleta e seleção das fontes quanto sua análise. Diante disso, é fundamental que o método histórico leve em consideração a aplicação de práticas de heurística digital coerentes tanto com as características das ferramentas e métodos utilizados, das fontes e dados trabalhados quanto com as reflexões teóricas básicas da disciplina histórica.

Nesse sentido, o presente projeto pretende avaliar de forma crítica os repositórios de periódicos digitalizados em língua portuguesa e suas interfaces gráficas, sobretudo àqueles ligados às Bibliotecas Nacionais brasileira e portuguesa e mapear repositórios similares que possuam acervos de periódicos de mais países dos PALOPs. A partir dessa avaliação, busco desenvolver um *framework* para pesquisa nos referidos repositórios, buscando encadear um fluxo de trabalho que atenda os aspectos metodológicos específicos da disciplina História, contemplando desde a pesquisa, coleta, tratamento e análise dos dados.

O *framework* é um enquadramento dos processos de pesquisa que será composto por uma biblioteca de referências bibliográficas sobre o temas – seus aspectos técnicos e teóricos –, a lista de repositórios de periódicos históricos em português digitalizados, uma documentação acerca de cada um dos repositórios e suas interfaces gráficas (apresentando as principais características, parâmetros possíveis de busca, opções de acesso e resultados); um conjunto de *Jupyter Notebooks*<sup>2</sup> com ferramentas para registro metodológico das pesquisas e resultados; indicação e manual de uso de programas de Reconhecimento Ótico de Caracteres (OCR) e análise de layout (OLR).

## Revisão da literatura

Muitas ferramentas e métodos digitais são rapidamente repelidos em contextos acadêmicos, taxados como inviáveis seja por sua curva de aprendizado muito grande ou por sua opacidade técnica (NASCIMENTO, 2020). Entretanto, seus usos são cada vez mais comuns e mesmo imprescindíveis para a execução de pesquisas nas ciências humanas. Desde o uso do e-mail institucional até algoritmos que

---

<sup>2</sup> Jupyter Notebook é um ambiente interativo que possibilita a criação de documentos que integram ao mesmo tempo códigos e textos. Um arquivo Jupyter Notebook, que possui a extensão `.ipynb`, permite que códigos sejam executados de modo interativo em diversas linguagens de programação, assim como é possível inserir textos com a sintaxe `Markdown`, sendo possível incluir imagens, links, vídeos, GIFs, entre outros elementos. Os resultados são apresentados no interior do próprio notebook, o que facilita seu uso para usuários sem domínio de linguagens de programação. Para mais informações, ver: <https://www.jupyter.org/>

processam linguagem natural, passando por editores de textos e planilhas, banco de dados, mapas, etc. Programas de computador, que nada mais são que uma longa sequência de códigos, caminham lado a lado com nossas pesquisas, aulas e publicações.

Contudo, como afirmam Röhle e Rieder, mesmo que parte significativa de nossas horas de trabalho sejam efetivadas na frente de uma tela, muitas das ferramentas que utilizamos não são enquadradas como parte de nossos métodos de pesquisa. Os autores entendem que a “função heurística dos métodos digitais de pesquisa” está focada em encontrar padrões, dinâmicas e relações nos dados, e que, mesmo que não estejam muito visíveis na cadeia metodológica, possuem repercussões epistemológicas significativas: computadores como ferramentas heurísticas “oferecem-nos perspectivas particulares sobre os fenômenos nos quais estamos interessados.” (RÖHLE; RIEDER, 2012, p. 70)

Portanto, como venho argumentando em trabalhos recentes (BRASIL e NASCIMENTO, 2020), a pesquisa em periódicos digitalizados também precisa estar sob rigoroso cuidado metodológico, que leve em consideração os variados e complexos aspectos que compõem tanto a pesquisa histórica quanto os elementos digitais que o formam.

Um jornal digitalizado, nas palavras de Maud Ehrmann e colaboradores, “é um objeto complexo determinado por múltiplas camadas de processamento e *dataficação*”. As interfaces dos repositórios digitais, portanto, têm um papel central na relação do usuário com a fonte: “Não apenas elas controlam o que os usuários podem aprender sobre o conteúdo digitalizado; elas também moldam ativamente os fluxos de trabalho do usuário, oferecendo diferentes seleções de ferramentas e recursos para pesquisar e explorar esse conteúdo” (EHRMANN; BUNOUT; DÜRING, 2019, p. 2). Ao mesmo tempo, comumente os usuários dessas interfaces “não estão conscientes dos vieses nos resultados de buscas, causados pelo processamento e *dataficação* dos jornais”. (PFANZELTER et al., 2020, p. 1).

Porém, é importante lembrar que esses pontos se aplicam a qualquer arquivo digital. Como argumenta Helle Jensen:

A literacia arquivística digital requer a compreensão de como a produção de arquivos digitais se baseia em designs técnicos que influenciam a sua usabilidade. Isso significa que (todos) os historiadores precisam adquirir competência digital em um nível profissional paralelo às habilidades que temos em compreender como a classificação e categorização de fontes afetam nossa interação com arquivos analógicos e moldam nossas questões de pesquisa. (JENSEN, 2021, p. 6, tradução minha)

A autora elenca três pontos fundamentais que devem ser encarados criticamente por historiadores/as ao lidar com arquivos digitais: a) categorias e etiquetas predefinidas; b) pesquisa por campos e organização de resultados; e c) metadados. As categorias e etiquetas estruturam o material arquivado de forma a “codificar” leituras e usos específicos; as pesquisas por campo e organização de resultados sugerem certos usos de seus recursos, mesmo se os usuários não precisarem seguir esses “usos

preferenciais”; e os metadados – dados sobre os dados –, que muitas vezes são o principal caminho para realizar buscas nos acervos e, portanto, seu registro – parcial ou incompleto – imputam certas interpretações às fontes. (JENSEN, 2021, p. 7–8)

Johan Jarlbrink e Pelle Snickars nos lembram, e deveríamos sempre levar isso em consideração ao clicar no botão de busca de qualquer acervo digital, que a “digitalização de um jornal histórico não é um processo neutro no qual os dados são transferidos de um meio a outro. Ao contrário, quando jornais são digitalizados eles são transformados.” E que, portanto, as maneiras como os dados digitais “são criados, armazenados, processados e formatados naturalmente tem implicações em como fontes históricas, como jornais digitais, podem ser acessados e usados, as histórias que podem ser exploradas e as histórias que podem ser (re) contadas” (JARLBRINK; SNICKARS, 2017, p. 1229–1230)

Para Salmi, o uso de interfaces gráficas, que possibilitem buscas e/ou acesso a grandes volumes de fontes, é essencial para os historiadores, entretanto,

para posterior análise digital do material é necessário ser capaz de fazer o download dos dados, por exemplo, como um dump de dados<sup>1</sup> que inclui os textos OCR e metadados, muitas vezes em padrões XML internacionais, como os formatos METS e ALTO (SALMI, 2020, posição 1350, tradução minha).

Portanto, inúmeras perguntas precisam ser encaradas ao lidar com esse conjunto de fontes: o acesso é feito através de interfaces gráficas que permitem busca por palavras? Quais formatos estão acessíveis? Existe OCR dos textos? Como os metadados são disponibilizados? Existem mais opções de visualização, tratamento e análise dos dados? Dependendo das respostas podemos avaliar até que ponto tais acervos podem ser acessados como Big Data e técnicas de leitura distante – distant reading (GOODING; TERRAS; WARWICK, 2013) – e Processamento de Linguagem Natural (PIROVANI; OLIVEIRA, 2018) podem ser aplicadas ou se se mantém fora de alcance no momento. Em outras palavras, os repositórios de periódicos históricos em português digitalizados permitem a implementação de técnicas e análises sofisticadas oriundas das Humanidades Digitais ou apenas emulam a pesquisa analógica?

Também é fundamental a efetivação de procedimentos de controle e registro metodológico, que muitas vezes são diluídos no manancial de dados disponíveis em acervos digitais. O registro desses procedimentos metodológicos são fundamentais para a pesquisa, visto que a menor alteração nos parâmetros de busca e procedimentos de pesquisa e coleta geraria resultados distintos (BRASIL; NASCIMENTO, 2020, p. 213).

Mats Fridlund e colaboradores chamaram atenção em obra recente:

Os métodos de pesquisa digital criam demandas novas e às vezes mais rigorosas de precisão, pensamento metodológico, auto-organização e colaboração do que a pesquisa histórica tradicional (FRIDLUND; OIVA; PAJU, 2020, posição 543, tradução minha).

Com o desenvolvimento de um *framework* para pesquisa histórica em periódicos digitalizados, busco contribuir para a elaboração de caminhos heurísticos digitalmente críticos e conscientes, entendendo que os métodos digitais de pesquisa, cada vez mais recorrentes, nos demandam esse cuidado. Esse objetivo torna-se ainda mais relevante ao olharmos para a atual literatura sobre o tema e constatarmos que parte significativa dos trabalhos versam sobre exemplos de hemerotecas que reúnem acervos em inglês, francês, alemão, finlandês, entre outros, mas com reduzida, ou nula, presença de acervos em português.

Em artigo publicado em 2019, Ehrmann, Bunout e Düring avaliaram vinte e quatro interfaces de hemerotecas digitais, sendo cinco dos Estados Unidos da América e dezenove europeias. Os autores desenvolveram um critério para avaliação das interfaces composto por seis pontos: crítica da fonte; busca de conteúdo, filtro de conteúdo, generosidade, gerenciamento e exploração do conteúdo pelo usuário, conectividade.

Após avaliação extensa, que envolveu *survey* com os pesquisadores que usaram as interfaces, os autores produziram um quadro com as principais características, possibilidades e limitações, dividindo tais interfaces em quatro gerações. Ao final, apontam que há uma “lacuna entre as expectativas crescentes do usuário, estimuladas pelos avanços da mineração de texto, e as capacidades atuais de interface.” (EHRMANN; BUNOUT; DÜRING, 2019, p. 17).

Entretanto, nenhuma das interfaces utilizadas conta com acervo em língua portuguesa, ou com hemerotecas digitais localizadas fora da Europa e Estados Unidos da América.

Eva Pfanzelter, Sarah Oberbichler, Jani Marjanen, Pierre-Carl Langlais, Stefan Hechl, avançaram em avaliação importante dos recursos e possibilidades de uso das interfaces de **ANNO** (biblioteca nacional da Áustria), **Digi and Korp** (biblioteca nacional da Finlândia e *The Language Bank of Finland*) e **Gallica** e **Retronews** (biblioteca nacional da França). Diferentemente do estudo de Ehrmann e colaboradores, aqui a metodologia envolveu a aplicação de estudos de caso, onde pesquisadores buscaram os temas ‘migração’, ‘nacionalismo’ e ‘gênero’ nas interfaces estudadas.

Para cada caso, criaram subcorpus para executar análises dos resultados. Esse ato de criar um corpus delimitado a partir de acervos digitais tem sido prática cada vez mais necessária na pesquisa na era digital. Segundo os autores, parte significativa das interfaces de hemerotecas digitais europeias cria essa demanda para usuários que pretendem utilizar seus acervos para além da lista simples de resultados.

Sem a possibilidade de criar subcorpora, os pesquisadores de humanidades muitas vezes não conseguem alinhar suas análises com suas questões de pesquisa específicas (PFANZELTER et al., 2020, p. 2).

Sua metodologia incluiu a avaliação desses corpus com as ferramentas *STM*, *TidySupervise* e *Overview*, criando assim um fluxo de trabalho para a análise de acervos digitais. Entretanto, mais uma vez tais

análises excluem os acervos em português.

Além das pesquisas aqui citadas, alguns projetos importantes tem sido desenvolvidos nos últimos anos sobre acesso, preservação e pesquisa em acervos de jornais digitalizados. Especificamente sobre jornais históricos, o projeto *impresso*: Media Monitoring of the Past<sup>3</sup> oferece um aplicativo on-line para pesquisa, análise e visualização de dados e textos minerados de mais de 200 anos de jornais europeus. O projeto, uma parceria do Laboratório de Humanidades Digitais do Instituto Federal Suíço de Tecnologia em Lausanne, com o Instituto de Linguística Computacional da Universidade de Zurique, e o Centro de História Contemporânea e Digital da Universidade de Luxemburgo, oferece recursos muito sofisticados de pesquisa e análise, com linhas do tempo, *ngrams*, tópicos, além da busca e acesso a metadados, texto gerado por OCR, criação de coleções pelo próprio usuário, recursos de comparação, entre outros.

### **Justificativa e relevância**

As pesquisas e projetos apresentados anteriormente reforçam a relevância historiográfica da pesquisa que imbrique as reflexões sobre teoria e metodologia da história com os aspectos técnicos e computacionais que engendram a existência de repositórios e interfaces gráficas para pesquisa em periódicos digitalizados. Ao mesmo tempo, apontam a necessidade da ampliação de investigações que incluam outras regiões do globo e outros idiomas. Aqui, pretendo abordar tais temas a partir de uma perspectiva conscientemente lusófona e partindo de um olhar do Sul Global. Pensar tanto a literatura sobre o tema, como apresentado aqui, quanto a própria existência ou não de recursos técnicos digitais nos países de língua portuguesa na América e África, em comparação com Portugal e o restante da Europa, se mostra um caminho analítico promissor para que possamos compreender o atual cenário da pesquisa histórica através de periódicos digitalizados e suas desigualdades globais.

O desenvolvimento de um *framework*, com ferramentas específicas em português para pesquisa em jornais escritos em português, é um resultado que visa mitigar tais desigualdades.

Além disso, o projeto pretende dar prosseguimento e aprofundar as pesquisas que venho desenvolvendo nos últimos anos. O trabalho desenvolvido em História Contemporânea e história Digital tem apresentado resultados relevantes para a academia brasileira, como pode ser visto na repercussão do artigo “História digital: reflexões a partir da Hemeroteca Digital Brasileira e do uso de CAQDAS na reelaboração da pesquisa histórica”, de autoria de Leonardo Fernandes Nascimento (UFBA) e meu, publicado na Revista Estudos Históricos, v. 33, n. 69, em 2020. Tal artigo conta com mais de 20 citações no *Google Scholar*.<sup>4</sup>

Tenho avançado nas reflexões dos aspectos técnicos referentes à interface da Hemeroteca Digital

---

<sup>3</sup>*impresso*: Media Monitoring of the Past, disponível em: <https://impresso-project.ch>.

<sup>4</sup>Citações no *Google Scholar* em 21 de fevereiro de 2022 nesse link

Brasileira (HDB), da Biblioteca Nacional. Como resultado, acabo de ter artigo aceito para publicação na Revista História da Historiografia, avaliação A1 na Capes, sobre uma ferramenta criada por mim, chamada [pyHDB](#), cujo objetivo é ser um auxiliar metodológico para os usuários da interface gráfica da HDB.

Atualmente desenvolvo o projeto “História Digital: acervos e ferramentas digitais para pesquisa e ensino” contemplado no edital PROPPG 04/2021, que já está em seu segundo ano. Na primeira parte da pesquisa analisamos as hemerotecas digitais da América Latina e Caribe, com trabalho da bolsista Ana Carolina Veloso. Atualmente analisamos a utilização de repositórios digitais por parte de docentes de cursos de história de Universidades localizadas na Bahia, com participação da bolsista PIBIC Priscila Valverde Silveira.

A realização do estágio pós-doutoral com esse projeto junto à Universidade Nova de Lisboa, sob supervisão do professor doutor Daniel Alves também consolida uma parceria internacional que já havia sido iniciada com o estabelecimento do protocolo de colaboração entre o LAB\_HD da Nova-Lisboa com o Laboratório de Humanidades Digitais da UFBA, ao qual estou também vinculado como pesquisador sênior. Essa parceria internacional possibilitou minha participação como editor do periódico *Programming Historian* em português, colocando um docente da UNILAB como colaborar numa rede internacional consolidada e cada vez mais relevante no campo da História.

Portanto, o estágio pós-doutoral trará contribuições significativas tanto para a minha formação acadêmica, quanto para o colegiado do curso de história e consequentemente o Instituto de Humanidades e Letras e a UNILAB, ao passo que visa desenvolver reflexões originais e ferramentas e métodos de pesquisa que atendam as demandas e interesses de docentes e discentes de nossa instituição e insere a UNILAB nessa rede internacional.

Esses resultados também são de interesse do Instituto de História Contemporânea da Universidade Nova de Lisboa, visto que abordam conjuntos de fontes, temáticas e regiões que dialogam estreitamente com linhas de pesquisa do instituto como “Colonialismo, Anti-Colonialismo e Pós-Colonialismo: Repensar Impérios e as suas Consequências” e “Usos do Passado: Memória e Patrimônio Cultural” e sobretudo com o Laboratório de Humanidades Digitais, que desde 2019 vem se constituindo como um espaço interdisciplinar para o desenvolvimento de pesquisas que conectam saberes e metodologias da área das Artes e das Humanidades com das Ciências Computacionais.

## **Objetivos**

### **Objetivos gerais**

1. Mapear acervos de periódicos digitalizados em português, com foco nas Bibliotecas Nacionais (Brasil, Portugal e demais PALOPs) e acervos de periódicos dos PALOPs digitalizados que se encontram em instituições variadas;



2. Avaliar os recursos, organização e estrutura dos acervos de jornais digitalizados em português;
3. Desenvolver um framework para pesquisa com periódicos históricos digitalizados em português;
4. Publicizar os resultados através da elaboração de lições para o Programming Historian em português e em artigos científicos

### Objetivos específicos

1. Analisar especificamente as interfaces gráficas dos repositórios;
2. Classificar as opções de acesso aos metadados e aos arquivos;
3. Identificar os impactos e relações entre a configuração de acesso e resultados gerados pelos repositórios para a pesquisa em História;
4. Analisar projetos relacionados à digitalização, busca e pesquisa em periódicos históricos na Europa
5. Avaliar ferramentas de OCR e OLR (Layout Parser; Kraken) específicas para periódicos;
6. Desenvolver ferramentas de registro metodológico e suporte para organização de data sets
7. Reunir os programas, scripts e ferramentas, devidamente comentadas e documentadas, para a criação do framework;

### Metodologia

- A primeira etapa da pesquisa engloba levantamento bibliográfico e organização de biblioteca de referências no Zotero;
- Em seguida, realizarei mapeamento dos repositórios e interfaces que serão utilizadas na pesquisa.
- Buscando avaliar tais repositórios e interfaces, será criado um formulário com os critérios de avaliação, inspirado em EHRMANN; BUNOUTtelle; DÜRING (2019).
- Será realizada análise crítica e técnica do projeto *impresso*.
- Avaliação das interfaces e repositórios
- Será feita pesquisa e avaliação de ferramentas de código aberto para Reconhecimento Ótico de Caracteres e automatização de reconhecimento de *layout* para documentos históricos.
- Criação de jupyter notebooks com passo a passo para criação de relatórios de registro metodológico;
- Criação de jupyter notebooks com scripts para organização e tabulação de dados e metadados;
- Todos os resultados serão organizados, documentados e comentados em repositórios públicos no *GitHub*
- Todas as etapas serão discutidas e apresentadas ao supervisor da pesquisa.
- Os resultados finais serão apresentados para os membros do LAB\_HD

## Resultados esperados

1. Biblioteca pública de referências bibliográficas utilizadas na pesquisa construída e disponibilizada no Zotero;
2. Framework para pesquisa com periódicos históricos digitalizados em português (reunindo *Jupyter Notebooks*, scripts, ferramentas e documentação);
3. Uma lição inédita para o Programming Historian
4. Produção de dois artigos científicos
5. Um *workshop* de formação para os membros do LAB\_HD

## Cronograma

Atividades	Out/22	Nov/22	Dez/22	Jan/23	Fev/23	Mar/23
Levantamento Bibliográfico e organização de biblioteca de referências no Zotero;	X					
Mapeamento dos repositórios e interfaces gráficas	X	X				
Criação do formulário com os critérios de avaliação		X				
Análise crítica e técnica do projeto <b>impresso</b>			X			
Avaliação das interfaces e repositórios			X	X	X	
Pesquisa e avaliação de ferramentas de código aberto para OCR e OLR					X	X
Criação de jupyter notebooks com passo a passo para criação de relatórios de registro metodológico						X
Escrita de artigo/lição				X	X	
Workshop para o LAB_HD						X

Atividades	Abr/23	Mai/23	Jun/23	Jul/23	Ago/23	Set/23
Pesquisa e avaliação de ferramentas de código aberto para OCR e OLR (CONTINUAÇÃO)	X					
Criação de jupyter notebooks com scripts para organização e tabulação de dados e metadados	X	X				
Escrita de artigo /lição			X	X		
Organização do framework e finalização do projeto				X	X	X

## Bibliografia

ALVES, Daniel. Ensinar Humanidades Digitais sem as Humanidades Digitais: um olhar a partir das licenciaturas em História. *Revista EducaOnline*, v. 15, n. 2, p. 5–26, 2021.

BRASIL, Eric; NASCIMENTO, Leonardo Fernandes. História digital: reflexões a partir da Hemeroteca Digital Brasileira e do uso de CAQDAS na reelaboração da pesquisa histórica. *Revista Estudos Históricos*, v. 33, n. 69, p. 196–219, 2020.

BRESCIANO, Juan Andrés. **La investigación histórica y las nuevas tecnologías**. Montevideo: Librería de la Facultad de Humanidades y Ciencias de la Educación, 2000. Disponível em: [https://www.academia.edu/2230307/La\\_investigaci%C3%B3n\\_hist%C3%B3rica\\_y\\_las\\_nuevas\\_tecnolog%C3%ADas](https://www.academia.edu/2230307/La_investigaci%C3%B3n_hist%C3%B3rica_y_las_nuevas_tecnolog%C3%ADas). Acesso em: 8 fev. 2020.

EHRMANN, Maud; BUNOUT, Estelle; DÜRING, Marten. Historical Newspaper User Interfaces: A Review. In: Athens, Greece, 2019. Disponível em: [https://www.researchgate.net/publication/339529189\\_Historical\\_Newspaper\\_User\\_Interfaces\\_A\\_Review](https://www.researchgate.net/publication/339529189_Historical_Newspaper_User_Interfaces_A_Review).

FICKERS, Andreas; CLAVERT, Frédéric. On pyramids, prisms, and scalable reading. *Journal of Digital history*, n. jd001, 2021. Disponível em: <https://www.journalofdigitalhistory.org/en/article/jXupS3QAeNgb>. Acesso em: 7 fev. 2022.

FRIDLUND, Mats; OIVA, Mila; PAJU, Petri (Orgs.). **Digital Histories: Emergent Approaches within the New Digital History**. Helsinki: Helsinki University Press, 2020. Disponível em: <https://hup.fi/site/books/e/10.33134/HUP-5/>. Acesso em: 5 jan. 2021.

GOODING, Paul; TERRAS, Melissa; WARWICK, Claire. **The myth of the new: mass digitization, distant reading, and the future of the book**. *Literary and Linguistic Computing*, v. 28, n. 4, p. 629–639, 2013.

JARLBRINK, Johan; SNICKARS, Pelle. Cultural heritage as digital noise: nineteenth century newspapers in the digital archive. *Journal of Documentation*, v. 73, n. 6, p. 1228–1243, 2017.

JENSEN, Helle Strandgaard. Digital Archival Literacy for (all) Historians. *Media History*, v. 27, n. 2, p. 251–265, 2021.

KEMMAN, Max. **Trading Zones of Digital History**. Berlin: De Gruyter Oldenbourg, 2021. Disponível em: <https://www.degruyter.com/document/doi/10.1515/9783110682106/html>. Acesso em: 23 set. 2021.

NASCIMENTO, Leonardo F. **Sociologia digital**: uma breve introdução. Salvador: EDUFBA, 2020. Disponível em: <http://repositorio.ufba.br/ri/handle/ri/32746>. Acesso em: 10 maio 2021.

NICHOLSON, Bob. The Digital Turn. *Media History*, v. 19, n. 1, p. 59–73, 2013.

PFANZELTER, Eva; OBERBICHLER, Sarah; MARJANEN, Jani; et al. Digital interfaces of historical newspapers: opportunities, restrictions and recommendations. *Journal of Data Mining and Digital Humanities*, 2020. Disponível em: [https://www.researchgate.net/publication/341926999\\_Digital\\_interfaces\\_of\\_historical\\_newspapers\\_opportunities\\_restrictions\\_and\\_recommendations](https://www.researchgate.net/publication/341926999_Digital_interfaces_of_historical_newspapers_opportunities_restrictions_and_recommendations).

RÖHLE, Bernhard Rieder Theo; RIEDER, Bernhard. Digital Methods: Five Challenges. In: BERRY, David M. (Org.). **Understanding Digital Humanities**. London: Palgrave Macmillan UK, 2012, p. 67–84. Disponível em: [http://link.springer.com/10.1057/9780230371934\\_4](http://link.springer.com/10.1057/9780230371934_4). Acesso em: 8 jun. 2021.

ROMEIN, C. Annemieke; KEMMAN, Max; BIRKHOLZ, Julie M.; et al. State of the Field: Digital History. *History*, v. 105, n. 365, p. 291–312, 2020.

SALMI, Hannu. **What is Digital History?** 1ª edição. Cambridge: Polity, 2020. (What is History? Series).

SOUTHERTON, Clare. Datafication. In: SCHINTLER, Laurie A.; MCNEELY, Connie L. (Orgs.). **Encyclopedia of Big Data**. Cham: Springer International Publishing, 2020, p. 1–4. Disponível em: [http://link.springer.com/10.1007/978-3-319-32001-4\\_332-1](http://link.springer.com/10.1007/978-3-319-32001-4_332-1). Acesso em: 16 nov. 2021.