



MECANISMOS DE BUSCA EM HEMEROTECAS DIGITAIS NACIONAIS: POSSIBILIDADES E DESAFIOS PARA A PESQUISA HISTÓRICA

Leandro Antonio de Almeida*

Universidade Federal do Recôncavo da Bahia – UFRB

leandroaalmeida@hotmail.com

RESUMO: O objetivo deste artigo é refletir sobre as possibilidades abertas para a pesquisa histórica pelos mecanismos de busca nas hemerotecas digitais brasileiras, avaliando um processo em curso do ponto de vista de um historiador e usuário dessas hemerotecas. Após tratar da relação entre informática e história desde os anos 1960, vamos analisar os mecanismos de busca da Hemeroteca Digital Nacional, Acervo Folha, Acervo Estadão, Acervo Veja e Diários Oficiais. A metodologia de análise parte da busca, nestas hemerotecas digitais, por palavras referentes a uma pesquisa de doutorado em curso sobre um desconhecido escritor brasileiro de pseudônimo João de Minas. Os mecanismos de busca acrescentaram relevantes informações à pesquisa, mas cuidados precisam ser tomados quanto à imprecisão dos resultados em função dos próprios mecanismos e da ortografia antiga.

PALAVRAS-CHAVE: Hemerotecas Digitais – Pesquisa Histórica – Mecanismos de busca – Jornais.

ABSTRACT: The main goal of this paper is to discuss the possibilities open for historical research by search engines in Brazilian digital newspaper archives, evaluating an ongoing process from the perspective of a historian and user of these newspaper archive. After dealing with the relationship between informatics and history since 60's, we will analyze the search engines of the Hemeroteca Nacional Digital (National Digital Newspaper Library), Acervo Folha de São Paulo, Acervo Estadão, Acervo Veja and Diários Oficiais (Official Gazettes Collection). The methodology of analysis is based on search, in these digital newspapers archive, for keywords referring to an ongoing doctoral research on an unknown Brazilian writer, João de Minas (nickname). The search engines have added relevant information to the research, but care must be taken regarding the inaccuracy of the results according to the mechanisms themselves and the old spelling.

KEYWORDS: Digital Newspapers Archive – Historical Research – Search Engine – Newspapers.

As tecnologias da informática modificam o modo de trabalho em inúmeras profissões, entre as quais a de historiador, auxiliando o armazenamento, consulta, análise e simulações das fontes do passado. Esta relação depende tanto das modificações por que passam essas tecnologias quanto das tendências internas ao

* Doutor em História Social pela Universidade de São Paulo. Professor da Universidade Federal do Recôncavo da Bahia.

campo historiográfico, que orientam os modos de apropriação abertos pelas potencialidades da informática.

Nos anos 1960, a preocupação com os elementos estruturais e coletivos das sociedades passadas levava a um uso das séries documentais, das quais eram recuperados os aspectos quantitativos, tarefa facilitada pela introdução, no campo da história, dos computadores então situados nos institutos de pesquisa. Segundo o depoimento de um influente historiador sobre a “revolução tecnológica” que assistiu, os historiadores foram “chamados pelo interesse da nova história pelos grandes números, postulado pela utilização de documentos que atingem as massas, como registros paroquiais na França, base da nova demografia, tornado necessário pelo desenvolvimento da nova história serial, o computador entrou na aparelhagem do historiador”.¹

Outra dessas revoluções da informática na historiografia foi notada na década de 1980, com o barateamento dos computadores pessoais, o desenvolvimento de interfaces gráficas, a integração entre objetos (textos, gráficos, tabelas, imagens etc.) e, sobretudo, um aumento da capacidade de armazenamento de dados, que fizeram as relações do historiador com o microcomputador evoluírem do mero processador de texto para programas de bancos de dados, gráficos, estatísticas e redes de comunicação. Luciano Figueiredo procurou refletir sobre o resultado desse desenvolvimento informacional até metade dos anos 1990 para o trabalho do professor de história e, principalmente, do pesquisador do passado. Mapeou as experiências existentes até o término da redação do seu texto, como a “catalogação, referencia e controle bibliográfico; editores ou processadores de texto; transcrição de dados e uso de scanners; planilhas, gráficos e tabelas; banco de dados; as redes de comunicação nacionais e internacionais”, além da utilização de softwares específicos para a atividade acadêmica e/ou historiográfica.²

Algumas dessas possibilidades apontadas pelo autor hoje se tornaram datadas ou tão comuns que dificilmente mereceriam comentários detidos, como a possibilidade de uso dos processadores de texto para transcrição de documentos em arquivos ou

¹ LEGOFF, Jacques. História. In: _____. **História e Memória**. Campinas: Unicamp, 1990, p. 135.

² FIGUEIREDO, Luciano R. História e informática: o uso do computador. In CARDOSO, Ciro Flamarion; VAINFAS, Ronaldo (Org.) **Domínios da História: ensaios de teoria e Metodologia**. Rio de Janeiro : Campus. 1997, p. 419-439.

redação de trabalhos científicos. Outras se desenvolveram de tal forma que se tornaram o foco de reflexão sobre a prática historiográfica nos anos 2000 e início de 2010. Nas palavras de um eminente historiador francês da atualidade, estudioso de uma das profícuas tendências do campo – a história da leitura e seus suportes materiais, “outra questão do nosso presente, menos aguda há dez anos, é a das mutações que impõe à história o ingresso na era da textualidade eletrônica”. Para ele, essa “era” abre novas possibilidades de construção, publicação e recepção dos discursos históricos, com impactos epistemológicos e cognitivos de vulto. O historiador não precisa mais se pautar pela lógica linear e dedutiva do texto impresso, podendo se permitir “uma articulação aberta, fragmentada, relacional do raciocínio, tornada possível pela multiplicação das ligações hipertextuais”. Longe da crença que deveria ter nas notas de rodapé, as quais remetem às fontes e locais de depósito dos documentos, hoje o leitor pode validar o argumento ao “se apoiar na consulta de textos (mas também de imagens fixas ou móveis, palavras gravadas ou composições musicais) que são o próprio objeto de estudo, com a condição de que, obviamente, sejam acessíveis em forma digital”.³

Claramente o pressuposto desta argumentação é o desenvolvimento vertiginoso da internet nos anos 2000 e o paralelo aumento de armazenamento e da velocidade de transmissão e recepção de dados. Isto possibilitou a viabilidade do armazenamento e disponibilização on-line de maior variedade de documentos históricos em diversos formatos como manuscritos, impressos, imagens, vídeos, voz, músicas, sem falar nas representações tridimensionais de objetos e espaços arquitetônicos, ensejando não só as visitas a “museus virtuais” como a simulação de paisagens e construções do passado.

Essas tecnologias digitais modificaram a relação do historiador com a versão digitalizada e digital do documento, pois se estabeleceu mecanismos de interação além do mero acesso. Uma dessas possibilidades, análogas àquelas abertas pelos filtros das bases de dados nos anos 1960, se abriu com o desenvolvimento de mecanismos de consulta conjugados a essas séries, os quais potencializam uma leitura de viés enciclopédico própria da textualidade eletrônica, feita na internet através dos hiperlinks. Estendendo as considerações de Chartier sobre os periódicos acadêmicos, a leitura emergente de periódicos antigos difere daquela feita no impresso porque nesta

³ CHARTIER, Roger. A história na Era Digital. In: _____. **A História ou a leitura do tempo**. Belo Horizonte: Autêntica, 2009 (2007), p. 59-60. Ele retoma a discussão feita em CHARTIER, Roger. Línguas e leituras no mundo digital. In: **Os desafios da Escrita**. São Paulo: UNESP, 2002, cap. 1.

a construção do sentido de cada artigo particular depende, mesmo inconscientemente, da sua relação com outros textos, que o antecedem ou o seguem e que foram reunidos dentro de um mesmo objeto impresso com uma intenção editorial imediatamente perceptível. A segunda leitura [eletrônica] age (...) a partir de uma organização enciclopédica do saber, que propõe ao leitor textos sem outro contexto além daquele de pertencer à mesma temática.⁴

Neste artigo vamos analisar essa relação entre a digitalização das massas documentais e possibilidades de busca de informação para pesquisa histórica a partir das consultas às hemerotecas digitais do Brasil. Para abordar esse problema, vamos partir dos impactos que os mecanismos de busca on-line tiveram em nossa pesquisa de doutorado, em curso na Pós-Graduação em História Social pela Universidade de São Paulo.

Nosso objetivo é estudar a atuação cultural de um escritor hoje pouco conhecido, Ariosto Palombo, jornalista que nas décadas de 1920 e 1930 foi mais conhecido pelo pseudônimo João de Minas, com o qual assinou inúmeros textos na imprensa brasileira e os livros que lançou. Nasceu em Ouro Preto por volta de 1896 e a partir de 1915 tornou-se revisor de **O Minas Gerais**, diário oficial da capital mineira. Em 1920, mudou-se para Uberaba onde colaborou no **Lavoura e Comércio** até o final da década. A partir de julho 1927 passou a escrever dominicalmente para o carioca **O Paiz**, enviando seus artigos de Uberaba. Pela editora desse jornal, em 1929, lançou seu primeiro livro **Jantando um Defunto**. Em setembro, foi a vez de o jornal paulista **O Correio Paulistano** passar a receber os seus textos. João de Minas fez campanha ao lado dos paulistas para as eleições presidenciais de 1930 como membro da Concentração Conservadora, partido político mineiro que rompeu com o presidente Antonio Carlos e com o Partido Republicano Mineiro quando se formou a Aliança Liberal pró-Getúlio. Talvez em função disso ganhou uma editoria política no jornal **O Paiz**, e mudou-se para o Rio de Janeiro em 30, lançando mais dois outros livros intitulados **Farras com o Demônio** e **Sangue de Ilusões**.

Com o movimento de outubro que destituiu Washington Luís, João de Minas fugiu para Uberaba e daí para a Argentina. Retornando, estabeleceu-se em São Paulo, primeiramente nas cidades de Franca e Araraquara em 1932, e em 1933 veio para a capital trabalhar no gabinete de seu amigo Dirlemando de Assis, então secretário de

⁴ CHARTIER, Roger. Línguas e leituras no mundo digital. In: _____. **Os desafios da Escrita**. São Paulo: UNESP, 2002, p. 28.

Viação e Obras Públicas da interventoria de Waldomiro de Lima. Colaborou por dois meses (junho e agosto) no **Jornal do Estado**, o diário oficial, até o fim da administração de Waldomiro, em agosto. A partir daí reorientou sua carreira, dedicando-se à ficção de massa. Recuperou seus escritos sertanistas, mixando e reeditando seus dois primeiros livros sob os títulos **Mulheres e Monstros** (1933) e **Pelas Terras Perdidas** (1934) e lançando um novo, o **Horrores e Mistérios nos Sertões Desconhecidos** (1934). Mas o forte de sua produção dessa fase foram os livros urbanos da coleção “Revolução Sexual Brasileira”, inspirados em Benjamin Costallat, cujos títulos revelam o teor: **A Datilógrafa Loura** (1934), **A Mulher Carioca aos 22 Anos** (1934), **Uma Mulher...Mulher!** (1934), **Fêmeas e Santas** (1935) e **A Prostituta do Céu** (1935). Em 1936 adentrou o gênero policial, publicando **Nos Misteriosos Subterrâneos de São Paulo**. Despediu-se da literatura com uma segunda edição de **A Mulher Carioca aos 22 Anos**, em 1937, pois estabeleceu uma seita religiosa sucessivamente nomeada Academia de Ciências Ocultas (1935), Igreja Brasileira Cristã Científica (1940) e Cura Divina Total (1959), com doutrina eclética de matiz nacionalista fundada no catolicismo popular, no espiritismo, na umbanda e no esoterismo. Adotando um novo pseudônimo, Mahatma Patiala, sua atividade na igreja ocupou seus esforços até pelo menos 1969 e produziu quatro edições de uma bíblia cujo primeiro volume, lançado em 1957, intitula-se **A Vida Começa na Ciência Divina**. João de Minas morreu em Boituva em 1984.⁵

Além da leitura da bibliografia especializada, o estudo sobre esse autor comporta uma dupla investigação documental. Uma vertente da pesquisa direcionou-se aos textos do escritor João de Minas publicados em livros, jornais e revistas ilustradas, nos meios de comunicação em que atuou no século XX. Ela permite entender as posições intelectuais, políticas e literárias que ele adotou em momentos distintos, reconstituindo sua biografia. A segunda vertente procurou informações de terceiros sobre o escritor, em especial textos opinativos sobre sua produção narrativa, que serviram como parâmetro de sua imagem entre os leitores e intelectuais brasileiros. Complementarmente, buscamos informações sobre as editoras e escritores que trataram de sua obra ou sua pessoa. A procura por fontes sobre a atuação do escritor mineiro se

⁵ ALMEIDA, Leandro Antonio de. Dos sertões desconhecidos às cidades corrompidas: um estudo sobre a obra de João de Minas (1929-1936). Dissertação (Mestrado em História Social), Universidade de São Paulo, 2008. Disponível em <http://www.teses.usp.br/teses/disponiveis/8/8138/tde-08072008-142247/> Acesso: 20/07/2012.

operou de duas maneiras. Na primeira, a partir das informações trazidas por seus livros e artigos, consultamos jornais e revistas ilustradas brasileiras. O mesmo valeu para informações sobre o escritor, cujas buscas concentraram-se em torno do lançamento de suas obras (1929-1937), focando as seções literárias ou de livros novos, além das notas de lançamento e propagandas.

A principal dificuldade dessa busca de informações em jornais e revistas depositados em arquivos é o volume. Às vezes pesquisávamos, impressos ou microfimes, um volume imenso da série periódica (por exemplo um ano ou mais) para encontrar alguma informação sobre o escritor mineiro. As mesmas dificuldades ocorrem em periódicos digitalizados que não dispõem de mecanismos de busca, a exemplo do Jornal do Brasil, digitalizado pelo Google, dos jornais digitalizados pelo Arquivo Público Mineiro,⁶ ou do jornal Última Hora, digitalizado pelo Arquivo do Estado de São Paulo.⁷ Na prática apenas disponibilizam acesso pela data de publicação, simulando online o ambiente de consulta nos arquivos. Estes processos ocorrem na maior parte das iniciativas de digitalização de hemerotecas.⁸

Além de possibilitar o acesso ao acervo por data/edição, algumas das recentes digitalizações de periódicos buscaram otimizar o processo de consulta com a incorporação de mecanismos de busca por palavras-chave, possíveis pelo trabalho prévio de transformação de imagem em texto, com softwares de tipo OCR. Foi o caso do Acervo Veja (1968-), disponibilizada pelo grupo Abril no final de 2008 / início de 2009; da página do Jus Brasil, dedicada aos Diários Oficiais de vários estados e do Distrito Federal desde o século XIX; do Acervo Folha, digitalizado em fevereiro de

⁶ Como o foco da coleção é o século XIX, seria muito difícil encontrar informações sobre João de Minas. Todavia, seu pai Pompílio Palombo viveu em Ouro Preto nessa época, o que abre perspectivas de encontrar informações sobre familiares. Há um mecanismo de busca no site, mas retorna os títulos dos jornais ao invés do conteúdo.

⁷ Esta instituição digitalizou exemplares de jornais e revistas ilustradas brasileiras. Ver o link Memória da Imprensa, disponível em <http://www.arquivoestado.sp.gov.br/memoriainprensa/>. Ver Jornal do Brasil (Rio de Janeiro). Google News. Disponível em. http://news.google.com/newspapers?nid=0qX8s2k1IRwC&dat=19360312&b_mode=2&hl=pt-BR ; Jornais Mineiros (Minas Gerais). Arquivo Público de Minas Gerais. Disponível em <http://www.siaapm.cultura.mg.gov.br/modules/jornais/search.php>; Última Hora (Rio de Janeiro). Arquivo do Estado de São Paulo. Disponível em <http://www.arquivoestado.sp.gov.br/uhdigital/>. Acesso: 27/07/2012. O acervo do Jornal do Brasil está sendo digitalizado pela Biblioteca Nacional, e vai compor em breve a Hemeroteca Digital Nacional.

⁸ Uma ampla listagem dessas iniciativas pode ser consultada nos sites da Biblioteca Nacional <http://www.bn.br/portal/index.jsp?nu_pagina=113>, Bibliotecno <http://bibliotecno.com.br/?page_id=493> e Biblioteca de Periódicos Digitais <<http://www.bibliotecno.com.br/periodicos/>>. Acesso: 27/07/2012.

2011, que abriga os jornais Folha da Manhã (1925-1959), Folha da Noite (1921-1959) e Folha de São Paulo (desde 1960); da Hemeroteca Digital abrigada na Biblioteca Nacional desde julho de 2012, com centenas de jornais e revistas brasileiras dos séculos XIX e XX que estão fora de circulação ou são raros; e do Acervo Estadão, do Estado de São Paulo, que em junho de 2012 lançou o acervo do seu jornal desde seu primeiro número (06/02/1875) até a atualidade.⁹ Não vamos estudar as interfaces e acessibilidades desses sites, devido ao curto espaço deste artigo e porque algumas análises já foram iniciadas, com amplos recursos visuais, pela série de postagens de Alex da Silveira no site Bibliotecno, dedicado às hemerotecas digitais.¹⁰ Vamos comentar os sites sob a ótica do historiador doutorando, ávido por encontrar informações sobre um escritor pouco conhecido.

A partir dessa necessidade, acessamos os sites das hemerotecas citadas acima e buscamos palavras-chave ligadas à nossa pesquisa, principalmente: João de Minas, Ariosto Palombo, Mahatma Patiala, os títulos dos livros e das editoras, e as variações da religião nomeada Igreja Brasileira Cristã Científica / Cura Divina Total / Academia de Ciências Ocultas. Tratamos apenas das informações novas à pesquisa, isto é, não contabilizamos os documentos desses acervos documentais que foram obtidos antes da consulta digital. Neste caso, o conhecimento da data e página foi suficiente para o acesso, prescindindo do mecanismo de busca. Certamente uma pesquisa nas etapas iniciais retornaria um número muito maior de ocorrências.

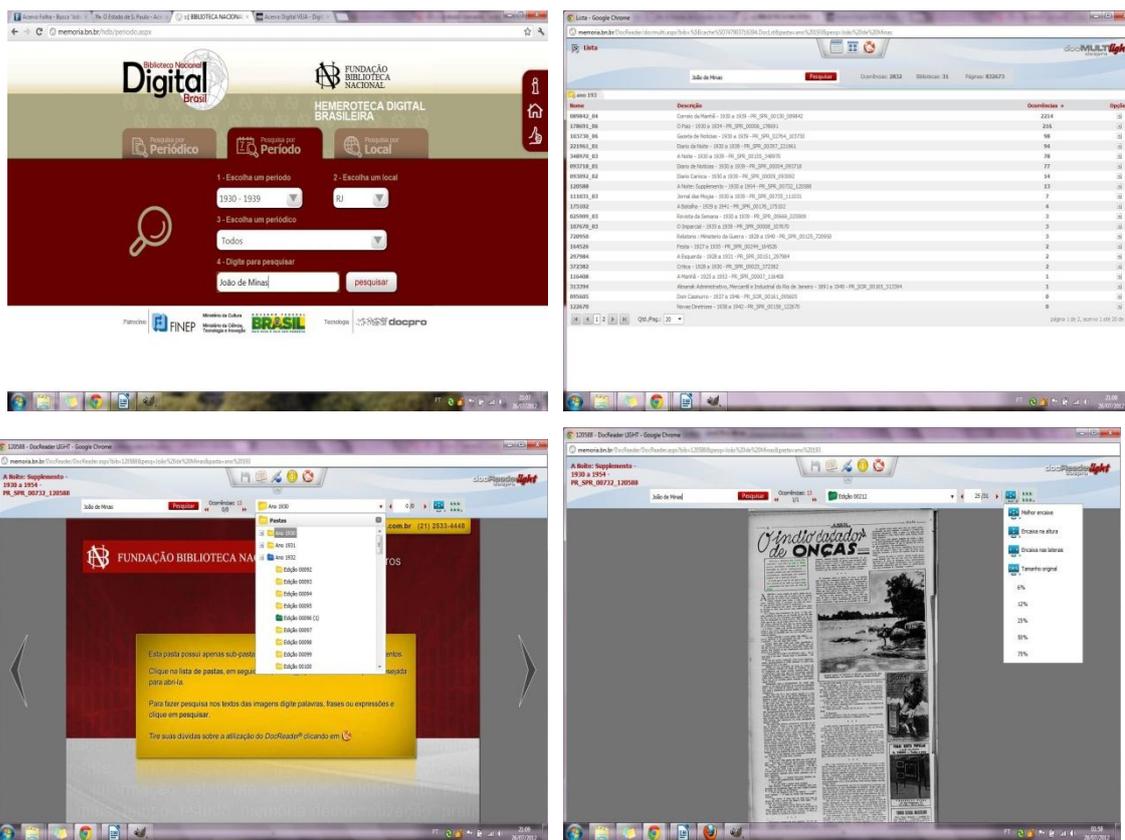
⁹ Ver: Acervo Folha (São Paulo). **Folha de São Paulo**. São Paulo. Disponível em <http://acervo.folha.com.br/>; Acervo Estadão (São Paulo). **O Estado de São Paulo**. São Paulo. Disponível em <http://acervo.estadao.com.br/>; Diários Oficiais. **Jus Brasil**. Disponível em <http://www.jusbrasil.com.br/diarios/>; Hemeroteca Digital Brasileira. **Fundação Biblioteca Nacional**. Rio de Janeiro. Disponível em <http://memoria.bn.br/hdb/periodico.aspx>; Veja Digital (São Paulo). Grupo Abril. Disponível em <http://veja.abril.com.br/acervodigital/>. Acessos em 27/07/2012.

¹⁰ Alex da Silveira é formado em biblioteconomia, especialista em sistemas de informação e funcionário da Biblioteca Nacional no Rio de Janeiro. Ver: SILVEIRA, Alex. Acervo histórico digitalizado do Jornal do Brasil no Google Notícias. Análise dos recursos de pesquisa. **Blog Bibliotecno**. Acesso em 18/04/10. Disponível em <http://bibliotecno.com.br/?p=594>; SILVEIRA, Alex. São necessários estudos de usabilidade das hemerotecas. **Blog Bibliotecno**. 08/09/10. Disponível em <http://bibliotecno.com.br/?p=1038>; SILVEIRA, Alex. Análise do acervo histórico digital da revista Veja disponível na web. **Blog Bibliotecno**. 22/04/12. Disponível em <http://bibliotecno.com.br/?p=618>; SILVEIRA, Alex. Mais um jornal disponível para acesso na web: O Estado de São Paulo. **Blog Bibliotecno**. 23/05/12. Disponível em <http://bibliotecno.com.br/?p=2223>; SILVEIRA, Alex. Biblioteca Nacional lança sua hemeroteca digital – parte 1. **Blog Bibliotecno**. 17/07/12. Disponível em <http://bibliotecno.com.br/?p=2270>; SILVEIRA, Alex. Biblioteca Nacional lança sua hemeroteca digital – parte 2. **Blog Bibliotecno**. 22 jul. 2012. Disponível em <http://bibliotecno.com.br/?p=2302>. Acessos: 22/07/12.

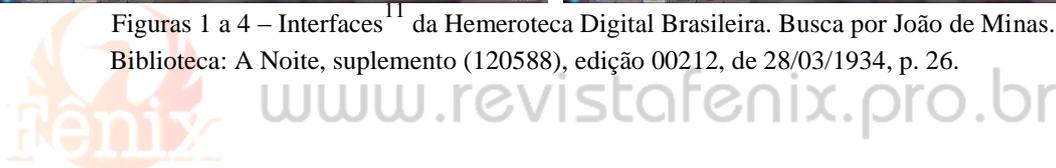
A pesquisa realizada entre fevereiro e julho de 2012 retornou 340 documentos com ocorrências sobre o escritor mineiro, entre colaborações, artigos que opinassem sobre sua pessoa ou seus livros, até simples menções do seu nome ou pseudônimo. A divisão por acervos retornou 45 documentos em Diários Oficiais (Jus Brasil) pertinentes à pesquisa, 35 no Acervo Folha, 246 na Hemeroteca Digital Brasileira, e 14 no Acervo Estadão. Apenas o Acervo da Veja não retornou nenhum resultado novo (o texto digital encontrado já estava em nossas mãos antes da consulta), muito provavelmente em função de este periódico ter sido criado numa época muito posterior à atuação do escritor mineiro.

Em função do porte, uma poderosa ferramenta de pesquisa é a busca por palavras da Hemeroteca Digital Brasileira, desenvolvida pela empresa DocPro e mantida pela Biblioteca Nacional. Seu principal destaque é a possibilidade de se consultar vários periódicos de um estado brasileiro em uma década, ou de consultar um termo em toda a série de um jornal ou revista específico, o que abre possibilidades de pesquisa histórica sincrônicas ou diacrônicas e também mostra ocorrências de um termo ou tema em periódicos de orientações políticas e finalidades sociais diversas. Os resultados retornados numa janela distinta da página de busca mostram a quantidade de ocorrências por “biblioteca” e, ao acessar uma delas, abre-se uma terceira janela com a série desejada (figuras 1 a 4).

Aí, no formato de pastas, há indicação, através de convenções de cores, das edições onde o termo aparece. Ao acessar uma dessas ocorrências, o termo buscado aparece destacado na página, em verde. As páginas dos periódicos podem ser visualizadas em diversos tamanhos, podendo ser também salvas no disco local, o que permite ao pesquisador montar no seu computador uma pequena biblioteca com as páginas que procura. Além disso, dois conjuntos de setas permitem navegação tanto entre as páginas da edição corrente como entre os resultados da busca dentro de uma biblioteca, este último de bastante valor para a pesquisa. Por exemplo, uma procura por “João de Minas” na biblioteca do jornal A Noite – Suplemento (Rio), entre 1930 e 1939, trouxe 13 ocorrências, que puderam ser consultadas diretamente na janela de visualização de página, sem retornar às buscas para seguir para a próxima edição onde o termo consta.



Figuras 1 a 4 – Interfaces¹¹ da Hemeroteca Digital Brasileira. Busca por João de Minas. Biblioteca: A Noite, suplemento (120588), edição 00212, de 28/03/1934, p. 26.



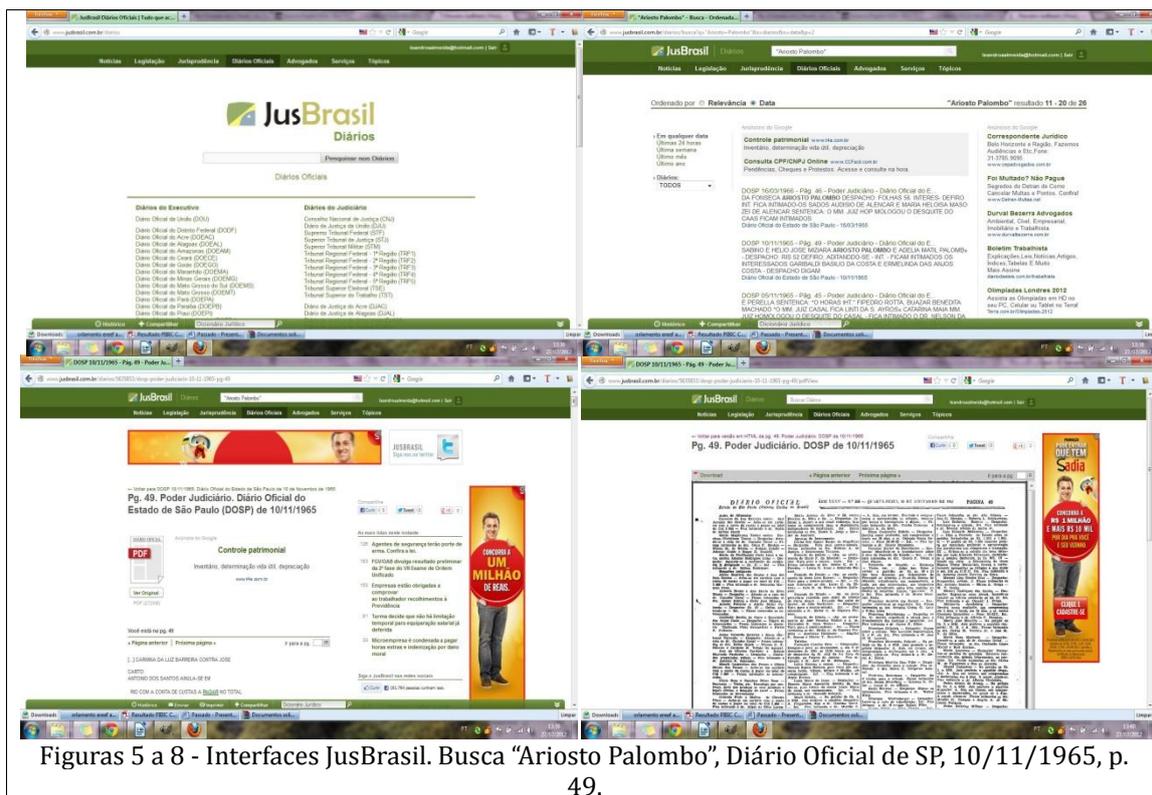
O principal inconveniente do mecanismo da Hemeroteca Digital Brasileira é a ausência da possibilidade de busca por termo exato. Para termos pouco comuns, o problema é minimizado, como na busca por “Ariosto Palombo”. No entanto, se o pesquisador digitar uma expressão comum na caixa de texto, poderá ter como retorno tanto a ocorrência exata como as palavras separadas na página, tornando a pesquisa inviável devido aos “ruídos” ou páginas não pertinentes. A título de exemplo, num jornal com alto volume textual como o **Correio da Manhã** (Rio) na década de 1930, a expressão “João de Minas” traz 2214 ocorrências, a grande maioria referentes à “Minas Gerais” e qualquer nome que contenha “João”.

Além desse inconveniente, no sentido oposto, seria interessante fortalecer os mecanismos de pesquisa textual em todo o acervo, não apenas em parte dele por vez. Atualmente, é preciso repetir uma pesquisa inúmeras vezes, variando as décadas e lugares, a não ser que se pesquise um único periódico. No nosso caso, a pesquisa

¹¹ Todas as imagens deste artigo foram capturadas da internet em 20 de julho de 2012.

“Ariosto Palombo” ou “João de Minas” teve que ser repetida inúmeras vezes, variando manualmente as décadas (1910-1919, 1920-1929, etc., até os anos 1980), cada uma destas variando os estados (RJ, SP, MG, etc.), para se encontrar resultados nos diversos periódicos selecionados com a opção “todos”. A repetição pode ser otimizada ao se integrar bibliotecas do acervo numa futura versão do programa.

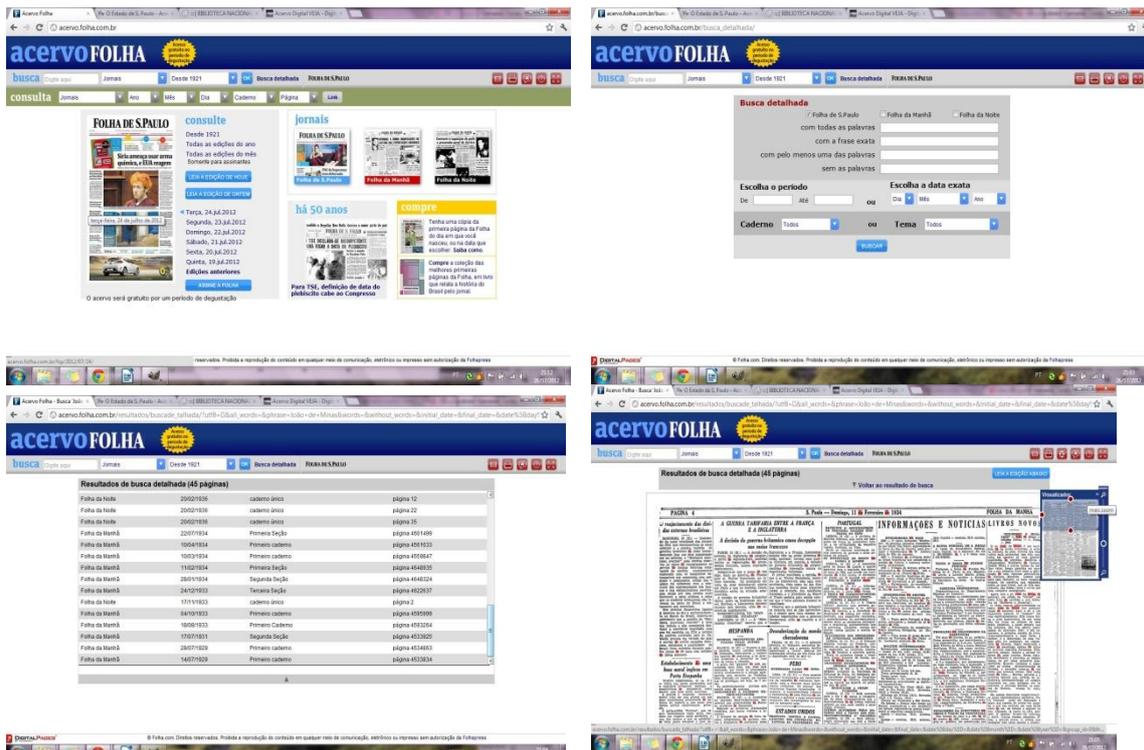
Muito provavelmente em função do menor porte e da tecnologia empregada, os outros acervos não apresentam essas limitações, pois permitem uma busca longitudinal por termos separados ou exatos, que podem ser escolhidos ao se optar pela pesquisa avançada, além da possibilidade de se personalizar com mais flexibilidade o período delimitado. O Acervo do JusBrasil com seus inúmeros diários oficiais, apresenta os resultados numa lista: valendo-se da tecnologia de buscas do Google, mostra o trecho com a ocorrência abaixo da referência. Ao se clicar no link, abre uma página com o conteúdo, mas pessimamente organizado para os Diários mais antigos. O mais viável é acessar o documento original em PDF num link disponível no canto superior esquerdo (figuras 5 a 8). Todavia, para localizar algo na massa textual deve-se abrir a caixa de busca e digitar o termo desejado, que fica destacado no centro de um quadrado azul. Apesar do escaneamento nem sempre estar bom, a opção de zoom permite ler o conteúdo, e outro botão permite salvar o PDF no disco.



Os três outros acervos tiveram suas interfaces desenvolvidas pela empresa Digital Pages, o que lhes dá certa semelhança. No caso do Acervo Folha, com três jornais diferentes, o inconveniente é que, para se buscar nos acervos, é preciso desmarcar as caixas de checagem, o que é pouco intuitivo.¹² Como aconteceu ao digitamos "João de Minas" na busca exata, se o pesquisador marcar todas as caixas, a pesquisa não retorna resultados. No caso onde todas estão desmarcadas, os resultados aparecem numa lista de links contendo as referências de título do jornal, data, caderno e página. Ao se clicar em um deles, a página do jornal aparece na tela, com os termos procurados destacados em vermelho, mas não é imediatamente possível lê-los, somente após acionar os quatro tamanhos de zoom disponíveis na barra à direita (figuras 9 a 12). Após a consulta, é possível voltar ao resultado da busca através de um link no centro da página. O inconveniente é a impossibilidade de salvar a página no disco, mas o pesquisador pode imprimir o arquivo e, aí, salvar em PDF.¹³ A resolução não é a mesma do zoom, o que dificulta a leitura.

¹² Essa informação foi obtida em consulta ao suporte do acervo digital da Folha.

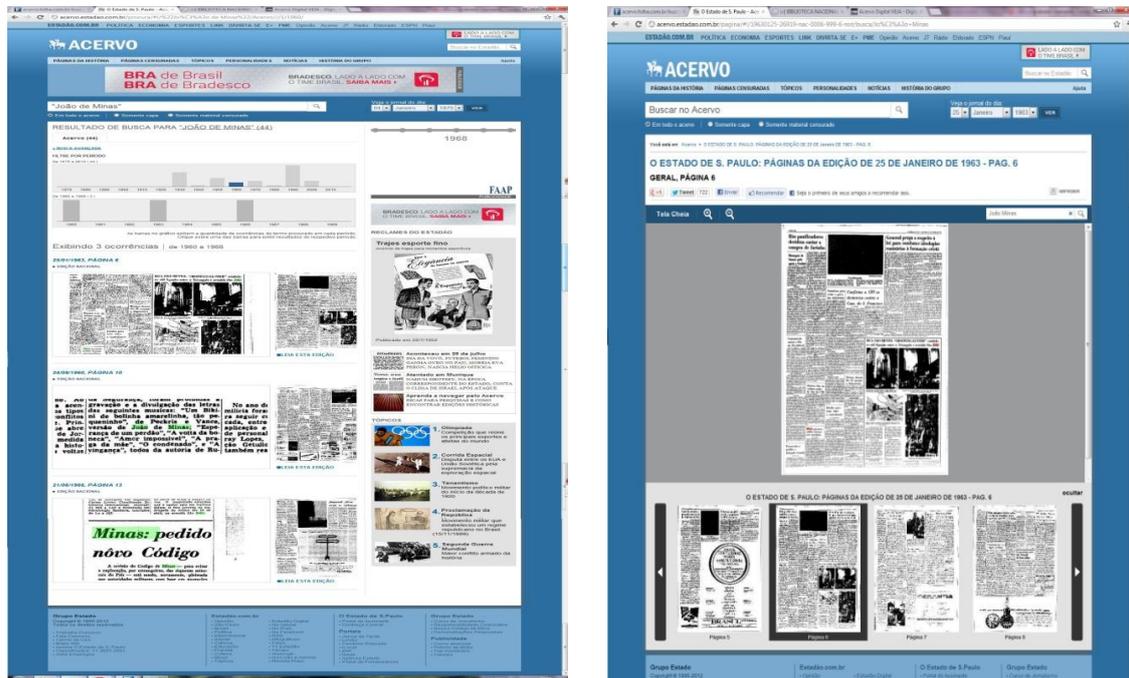
¹³ SILVEIRA, Alex. Mais um jornal disponível para acesso na web: O Estado de São Paulo. **Blog Bibliotecno**. 23/05/12. Disponível em <http://bibliotecno.com.br/?p=2223>. Acesso: 22/07/12.



Figuras 9 a 12 - Interfaces do Acervo Folha. Busca exata por João de Minas, pagina da Folha da Manhã de 11/2/1934

O Acervo Estadão tem uma boa navegação e filtros. Ao se digitar um termo na busca, logo abaixo da caixa aparecem gráficos em barra com os resultados por décadas. Ao se clicar em um deles, outro gráfico aparece logo abaixo, filtrando os resultados por anos. Depois, são visualizadas uma lista com as ocorrências, com as referências de data e página, seguido de uma prévia da seção, com o termo buscado destacado em verde e, à sua direita, visualiza-se uma miniatura da página inteira. Um click leva o pesquisador a outra janela, com a página inteira em tamanho maior, com botões de zoom que possibilitam visualizar a matéria em diversos tamanhos, além da possibilidade de se navegar nas páginas da edição através das miniaturas (figuras 13 e 14). O único inconveniente, já apontado pelos artigos Alex Silveira e Alice Melo,¹⁴ é que o acesso ilimitados às páginas de maior resolução é restrito a assinantes do jornal. Um cadastro pode ser feito gratuitamente, mas o número de visualizações totais restringe-se a 20 por mês, número ínfimo para uma pesquisa histórica de vulto. O acervo está todo disponível gratuitamente para universidades parceiras, mas o deslocamento a uma delas se torna inconveniente.

¹⁴ SILVEIRA, Alex. Mais um jornal disponível para acesso na web: O Estado de São Paulo. **Blog Bibliotecno**. 23/05/12. Disponível em <http://bibliotecno.com.br/?p=2223>. Acesso: 22/07/12; MELO, Alice. Memória Dinâmica. **Revista de História da Biblioteca Nacional**. 04/07/12. Disponível em <http://www.revistadehistoria.com.br/secao/reportagem/memoria-dinamica>. Acesso: 17/07/12.



Figuras 13 e 14 - Interfaces do Acervo Estado. Busca por “João de Minas”, anos 60, item 2 (25/01/1963)

Assim como aconteceu com o grupo Estado, no site da Folha é possível ver a mensagem de que o acesso irrestrito tem previsão para acabar, pois se trata de um “período de degustação”. Como destaca o pesquisador Ricardo Pimenta,¹⁵ o modelo do acervo Estado é um indício que “a memória se tornou um capital. Ela é importante hoje, pois há um consumo dela. Quem a controla, possui controle do discurso sobre seu próprio passado e reconhecimento no espaço público”, além dos lucros financeiros advindos desse interesse. É uma questão análoga ao que ocorre com a digitalização de livros pelo Google:¹⁶ como lidar com o fato de acervos de interesse público, mesmo que privativos, estarem sendo digitalizados por megacorporações.¹⁷ Ao contrário da

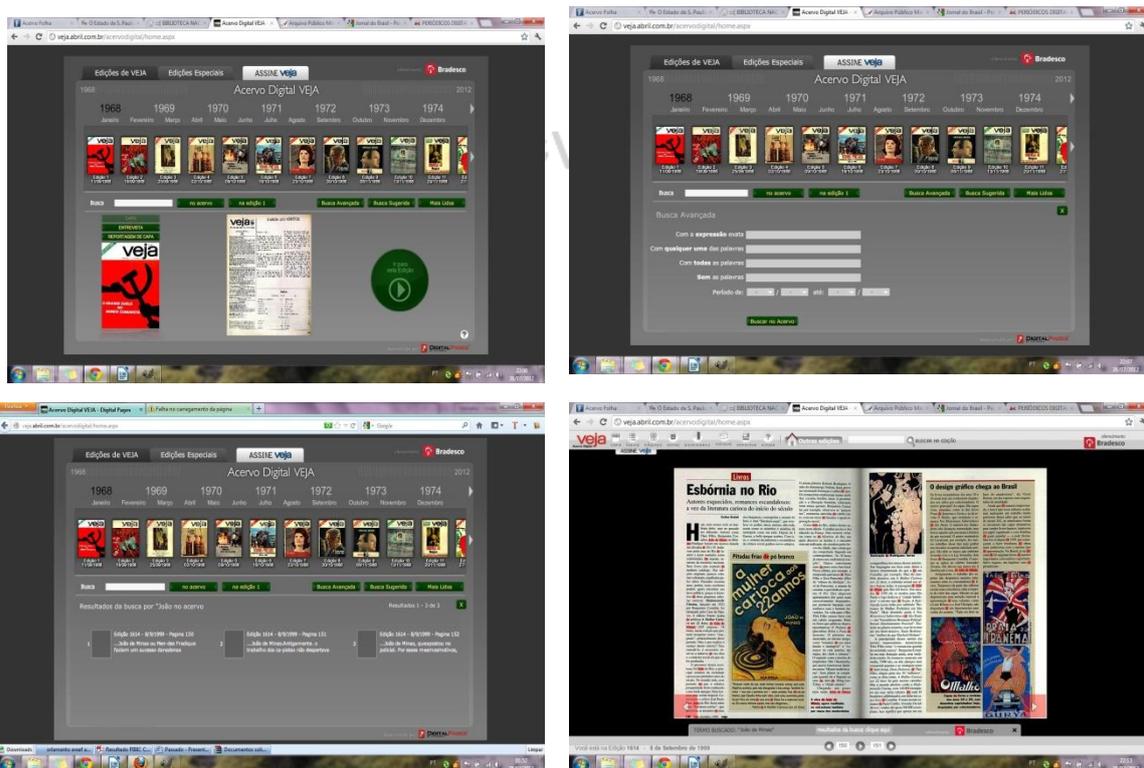
¹⁵ MELO, Alice. Memória Dinâmica. *Revista de História da Biblioteca Nacional*. 04/07/12. Disponível em <http://www.revistadehistoria.com.br/secao/reportagem/memoria-dinamica>. Acesso em: 17/07/12.

¹⁶ Do mesmo modo que as hemerotecas digitais, o Google Books que também é uma interessante ferramenta de pesquisa para historiadores, porque há um mecanismo de busca que traz referências ao termo em livros e revistas de diversos períodos. A depender do grau de liberação dos direitos negociados entre as editoras e a empresa, pode-se visualizar páginas inteiras onde os termos ocorrem, ou apenas algumas linhas, ou então a visualização pode estar indisponível. De qualquer forma, como há o retorno da referência, o exemplar físico pode ser consultado em uma biblioteca física. Em nossa pesquisa de doutorado, também utilizamos esse mecanismo para encontrar palavras-chave acima citadas, que acrescentaram importante volume de informações.

¹⁷ DARNTON, Robert. *A questão dos livros: passado, presente e futuro*. São Paulo: Cia das Letras, 2010.

Biblioteca Nacional, que equalizou a questão com financiamento público e digitalizou periódicos extintos, o grupo Veja contou com parceria de um banco privado, o Bradesco, que considerou viável o investimento de imagem da marca.¹⁸

As características do seu acervo, o fato de se tratar de uma revista de circulação relativamente recente, diminuem alguns problemas encontrados nas iniciativas com jornais, como a eficácia dos mecanismos de busca. Os resultados desta são apresentados em listas numeradas, contendo a referência de data, edição e páginas, além de uma linha onde o termo ocorre. Ao se clicar, abre-se uma página dupla da revista que, ao contrário das anteriores, é colorida (figuras 15 a 18). A visualização traz os termos buscados com destaques vermelho, e contém apenas um nível de zoom (aumenta ou diminui). Há também a opção de abrir as miniaturas das páginas da edição e de retornar aos resultados da busca, que aparecem em uma lista simples. Não há opção de salvar a página em disco, mas é possível imprimi-las, inserir notas em uma caixa texto e marcadores para os favoritos.



Figuras 15 a 18 - Interfaces do Acervo Veja Digital. Busca avançada por João de Minas, opção 2. Edição 1614, 08/09/1999, p. 152-3.

¹⁸ SILVEIRA, Alex. Mais um jornal disponível para acesso na web: O Estado de São Paulo. **Blog Bibliotecno**. 23/05/12. Disponível em <http://bibliotecno.com.br/?p=2223>. Acesso em 22/07/12.

Como pode ser percebido na figura, um dos grandes diferenciais do acervo da VEJA é que a tecnologia de digitalização permite uma alta qualidade na visualização de imagens, mantendo não apenas a sua cor como alta resolução. Vale para os quatro outros acervos o que Alex da Silveira já notou ao comparar o acervo mantido pela Abril com o Jornal do Brasil: “enquanto o Google disponibiliza os jornais em preto e branco devido ao fato da digitalização ter ocorrida a partir de microfimes, o acervo da Veja disponibiliza todas as suas páginas em cores”.¹⁹ A perda de cores e de qualidade na digitalização de imagens a partir de microfimes tem impacto da numa pesquisa histórica com foco na análise iconográfica, o que precisa ser levado em conta ao se propor digitalizações de periódicos. No nosso caso, encontramos em formato digital, tamanho grande, a capa colorida da rara primeira edição de **A Mulher Carioca aos 22 Anos**, um dos livros mais comentados de João de Minas.

Algumas considerações gerais sobre a eficácia da busca de informações histórica nos acervos de jornais precisam ser feitas. A primeira é a margem de imprecisão existente nos programas. Há grande possibilidade de o mecanismo de busca deixar de fora edições de jornais onde o termo ocorre, como aconteceu nos programas da Folha, Estado, JusBrasil e Biblioteca Nacional. Como havíamos feito pesquisas sobre João de Minas antes da digitalização, encontramos informações nos periódicos físicos que não apareciam nas buscas virtuais, e só puderam ser acessadas com a consulta à página via data. Foi o caso de parte da extensa colaboração de João de Minas no jornal carioca O Paiz (Rio) em 1930; do artigo de Plínio Barreto sobre o seu primeiro livro em 1929, publicado no Estado de São Paulo; de parte das colaborações do escritor mineiro no Diário Oficial do Estado de São Paulo em 1933; além da reportagem de Carlos Lacerda sobre a Igreja Cristã Científica para a Folha da Noite em 1959. Neste caso, o problema maior foi o péssimo estado do jornal escaneado, pois a referida reportagem está ilegível. Nos outros, inclusive na Folha, às vezes o programa confunde termos ao trocar “Unitas” (uma editora paulista da década de 1930) por “União”, ou desconsiderar hifenizações (a busca por ciência retorna consciência).

Essa imprecisão pode ocorrer por causa das grafias antigas. Em geral, os buscadores atualizam-na, mas há grande possibilidade de resultados ficarem de fora da busca. Para citar um exemplo, apesar de a Hemeroteca Digital Brasileira desconsiderar

¹⁹ SILVEIRA, Alex. Análise do acervo histórico digital da revista Veja disponível na web. **Blog Bibliotecno**. 22/04/12. Disponível em <http://bibliotecno.com.br/?p=618>. Acesso em 22/07/12.

as diferenças entre maiúsculas / minúsculas e acentuação, a busca no acervo do Correio Paulistano retorna resultados diferentes nas buscas “Ciencias Ocultas” (0 páginas) e “Sciencias Occultas” (57 páginas) ou “datilografa” (0 páginas) e “Dactylographa” (107 páginas). Nos outros acervos de jornais não foi diferente. Portanto, a questão da grafia precisa estar no campo de atenção quando o pesquisador do passado realizar buscas textuais nesses programas.

Uma questão de outra ordem é considerar aquilo que se perde na digitalização desses acervos. É preciso ter sempre em mente que há um processo de transposição da imagem do documento para outro meio, no qual fica de fora sua materialidade, como o tipo do papel empregado. Numa análise que tome o próprio periódico como objeto²⁰, essa é uma dimensão cada vez mais considerada, podendo prover diversas informações para os historiadores sobre o porte do jornal ou revista, tecnologias de impressão, etc. Daí que os investimentos em conservação de hemerotecas físicas e constituição de arquivos devem ser igualmente vultosos, evitando que a disponibilização de informações seja acompanhada da destruição acelerada da materialidade das fontes: “como leitores, como cidadãos, como herdeiros do passado, devemos, pois, exigir que as operações de digitalização não ocasionem o desaparecimento dos objetos originais e que seja sempre mantida a possibilidade de acesso aos textos tais como foram impressos e lidos em sua época”.²¹

Apesar das limitações, como vimos, os mecanismos de busca textual das hemerotecas digitais abrem possibilidades de pesquisa histórica até então indisponíveis. Nesse sentido, a capacidade de localização de informações nesses acervos digitais deu um salto análogo àquele aberto pelos processadores de cálculo nos anos 1960, os quais possibilitaram tratar grandes séries documentais em grande escala. Da mesma forma, os mecanismos de busca textual permitem abrir e incrementar pesquisas de modos até então impensáveis, a exemplo da que realizamos sobre um escritor pouco conhecido, sobre o qual há pouquíssimas informações sistematizadas nas enciclopédias e livros de

²⁰ Sobre a distinção operacional entre os jornais como fonte e como objeto de pesquisa histórica no Brasil, ver DE LUCA, Tânia. Fontes Impressas: história dos, nos e por meio dos periódicos. In: PINSKY, Carla Bassanezi (Org.). **Fontes Históricas**. São Paulo: Contexto, 2006, p. 111-153.

²¹ CHARTIER, Roger. Línguas e leituras no mundo digital. In: _____. **Os desafios da Escrita**. São Paulo: UNESP, 2002, p. 29. Essa questão da relação entre desenvolvimento das tecnologias da informação e destruição do passado foi desenvolvida de forma mais ampla em STILLE, Alexander. **A Destruição do Passado**: como o desenvolvimento pode ameaçar a história da humanidade. São Paulo: Arx, 2005.

literatura. Para desenvolver essas possibilidades, os problemas técnicos levantados podem ser facilmente corrigíveis em versões mais atualizadas e poderosas dos programas.

Mais complexa e longa é a questão da disponibilidade (gratuita ou paga) dos acervos analisados e daqueles em processo de digitalização, como **A Gazeta** (São Paulo), questão que envolve concepções sobre a finalidade dessas iniciativas e da função da memória na sociedade atual e futura. É possível que as iniciativas governamentais e institucionais possam orientá-las na perspectiva de que esses acervos se tornem disponíveis para a totalidade dos cidadãos, tornando essas fontes acessíveis a qualquer um que deseje conhecer e formular novas perguntas sobre a sociedade atual com base no passado.



www.revistafenix.pro.br