

DE POR QUE E COMO PRECISAMOS DE BIBLIOTECAS DIGITAISAlckmar Luiz dos SANTOS¹

Resumo: Neste texto discuto o estado atual da construção de bancos de dados e de bibliotecas digitais de literatura brasileira, tentando mapear alguns processos, estratégias e dificuldades, sejam da área de letras, das instituições governamentais ou dos próprios pesquisadores.

Palavras-chave: bibliotecas digitais, bancos de dados digitais, literatura brasileira

Abstract: In this paper I discuss the current construction state of database and digital libraries of Brazilian literature, trying to examine some processes, strategies and difficulties either in the Letters field, governmental institutions or in the researchers themselves.

Key-words: digital libraries, digital database, Brazilian literature.

Em comparação a literaturas de outras línguas, estamos muito atrás no que se refere à digitalização de obras literárias. Até mesmo literaturas como a latina e a grega se encontram na frente das de língua portuguesa. Ora, para que se viabilize a digitalização de obras literárias, antes dos necessários (e até modestos) investimentos financeiros, é necessário que se tome consciência da importância desse gênero de projeto. E, ainda, é preciso uma visão panorâmica do estado (problemático) em que já se encontram edições digitais disponíveis de obras literárias em língua portuguesa.

Algumas pesquisas no *Google* permitem uma boa idéia do que está disponível em termos de obras literárias digitalizadas. Primeiramente, podemos buscar por bibliotecas digitais, tentando encontrar resultados em Inglês, Francês, Espanhol e, claro, Português. Para “biblioteca digital”, encontramos **846.000** respostas²; “bibliothèque électronique” deu **231.000** resultados; “biblioteca electrónica”, **373.000**; “digital library” e “digital libraries”, **5.672.000**. Numa primeira visada, esses dados parecem contradizer parcialmente o que afirmamos acima.

Todavia, essa pesquisa não é confiável, na medida em que os resultados trazem, no mais das vezes, referências indiretas a bibliotecas digitais. Por exemplo, uma das respostas leva ao endereço www.cosmo.com.br/especial/cosmo_especial/2003/03/22/materia_esp_53366.shtml, que não leva a nenhuma biblioteca em meio eletrônico, mas a um comentário acerca da biblioteca digital de teses da Unicamp. Se queremos mesmo ir atrás de obras digitalizadas, é preciso fazer uma pesquisa mais direta, perguntando sobre carregamento (*download*) de obras. Assim, a busca por “Literary texts” “download”, nos dá **94.800** respostas;

por “textos literários” “download”, **13.900**; por “textos literarios” “download”, **521**; por “textes littéraires” “téléchargement”, **13.500**.

Mesmo aqui, é preciso analisar os dados que aparecem. Uma maior quantidade ainda não significa necessariamente que vamos encontrar mais obras digitalizadas. Se é esperado que tenhamos mais obras em Inglês do que em Português, não podemos supor que tenhamos mais em Português do que em Francês, apenas pela comparação entre as quantidades encontradas. É preciso também enveredar pelos sítios que oferecem as obras digitalizadas e analisar essa oferta, além da qualidade da digitalização. As 13.900 respostas que encontramos em Português, comparadas às 13.500, em Francês, podem significar simplesmente que há muitos sítios em Português oferecendo as mesmas obras (uns copiando dos outros, prática inerente à internete, mas exercitada com mais frequência por páginas em nosso idioma).

O projeto *Gutenberg*, que pretende ser uma espécie de Biblioteca de Alexandria da literatura universal, traz mais de 20.000 obras em formato digital, mas apenas 54 são de língua portuguesa. E, destas, somente 3 são brasileiras (se assim considerarmos Anchieta). Ainda em língua inglesa, o portal *Google* vem desenvolvendo seu projeto gigantesco de digitalização de mais de um milhão de obras (não apenas literárias), em conjunto com várias bibliotecas públicas de vários países. Apenas uma busca simples, pela palavra “Brasil”, no *GoogleBooks*, nos dá **461.000** respostas. Além disso, as estratégias de digitalização em imagem do Google permitem, com grau variável de confiabilidade, uma busca por palavras, mesmo sendo o documento em formato imagem (o que, sendo minimamente confiável, permitiria reduzir bastante o tempo e os custos de digitalização, pois não haveria mais a necessidade de passar todas as imagens para o formato texto).

Voltando, porém, a projetos focados especificamente em algumas literaturas nacionais, vamos ver que tanto as intenções quanto as realizações são bem mais modestas. Em Italiano, por exemplo, uma lista de sítios está disponível em www.librialice.it/virtual/net.vir/vnetita.htm. Dentre eles, a *Biblioteca dei Classici Italiani* (www.classicalitaliani.it/) dispõe de 453 obras digitalizadas. Já a *Biblioteca della Letteratura Italiana* (www.letteraturaitaliana.net/) traz 342 obras de 205 autores.

Em Espanhol, a *Biblioteca Virtual Miguel de Cervantes* (www.cervantesvirtual.com) afirma que sua “sección de Literatura presenta reunidas miles de obras digitalizadas en el ámbito de los diferentes géneros literarios y de la historia de la literatura. Novela, teatro, poesía y crítica literaria se dan la mano en este catálogo de la literatura española e iberoamericana que reúne obras de la Biblioteca Nacional de España, de la Real Academia de la Lengua, de otras Academias y Bibliotecas Nacionales de América, y de diversas editoriales. En este espacio se pueden encontrar también ediciones facsímiles, actas de congresos, revistas o manuscritos relacionados con la literatura hispánica e iberoamericana”. Ela traz obras digitalizadas diretamente em *html*, assim como em imagens e registros sonoros, não só em

Espanhol, mas também em Português. E, além das próprias obras, traz também uma ferramenta de concordância. No total, são 720 referências em literatura geral, sendo algumas poucas da literatura brasileira. Na Argentina, o jornal *El Clarín* mantém o projeto *Biblioteca Digital Argentina* (www.biblioteca.clarin.com/pbda/index.html), com 157 obras digitalizadas.

Em Francês, a *Bibliothèque Universelle* (abu.cnam.fr) contém 288 obras de 101 diferentes autores. A *Bibliothèque Electronique du Québec* (jydupuis.apinc.org/) publica escritores francófonos : 217 obras do Québec e 537 outras de diferentes países (na maioria, francesas). A *Bibliothèque électronique de Lisieux* (www.bmlisieux.com) apresenta 461 obras digitalizadas. Mas é sem dúvida *Gallica* (gallica.bnf.fr) o maior projeto de digitalização. São 4.000 obras em formato imagem e 1.250 em texto.

Chegando à língua portuguesa, a *Biblioteca Nacional de Lisboa* (bnd.bn.pt) põe à disposição dos leitores um total de 7104 “registos”, dos quais 358 são de edições digitais. Mas é conveniente esclarecer que se trata de obras digitalizadas em formato imagem, sem que se possa manipular a obra como se fosse texto digitado. No Brasil, houve há alguns anos a tentativa da *Biblioteca Nacional* (www.bn.br) de fazer a digitalização, em formato texto, de uma grande quantidade de obras. Essas excelentes intenções esbarraram em dificuldades e barreiras de política cultural, limitando-se a menos de 180. Mais recentemente, houve a iniciativa do Ministério da Educação de fazer um sítio com obras digitalizadas, no projeto *Domínio Público*. Todavia, o MEC não fez opção de investir num projeto de digitalização de obras literárias. Limitou-se a tomar emprestados arquivos de outros sítios, numa timidez incompatível com suas próprias atribuições de órgão governamental encarregado de zelar pela educação (o que ainda acarreta a consequência de se estar apenas reproduzindo atividades, acertos e equívocos de outros). O sítio *Prossiga* do CNPq (www.prossiga.br/pacc/bvl), embora anuncie uma *Biblioteca Virtual de Literatura*, não faz mais do que listar bibliotecas digitais do Brasil e do mundo. Na Universidade de São Paulo, a Escola do Futuro tem, há anos, seu projeto *Biblioteca Virtual do Estudante Brasileiro* (www.bibvirt.futuro.usp.br/index.php), com 214 títulos integrais das literaturas portuguesa e brasileira. O portal da Universidade de Campinas hospeda o sítio do projeto *Memória de Leitura* (www.unicamp.br/iel/memoria), que reúne mais de 220 títulos. Contudo, não se trata unicamente de obras integrais; em boa parte, são textos isolados como contos, poemas, crônicas etc. O sítio de literatura do *Núcleo de Pesquisas em Informática, Literatura e Lingüística* (www.literaturabrasileira.ufsc.br), além de dados de história literária (a respeito de mais de 14.500 autores e de mais de 54300 obras), está colocando ao alcance do internauta praticamente 500 títulos integralmente digitalizados, no que é talvez o maior projeto do gênero em literaturas de língua portuguesa. O *Jornal de Poesia* (www.secrel.com.br/jpoesia/poesia.html), heróica iniciativa do poeta cearense Soares Feitosa, é feito sem qualquer tipo de apoio e, mesmo não colocando o conteúdo integral das obras, é uma referência preciosa para quem busca poemas em língua portuguesa (há ali mais de 3.000

poetas e críticos da literatura, do passado e da atualidade, misturando clássicos a contemporâneos de interesse desigual). Há, ainda, uma certa quantidade de sítios que se limitam a copiar e a oferecer a seus leitores arquivos já digitalizados por esses acima mencionados. De toda maneira, se contamos tudo que se encontra disponível em termos de obra literária digitalizada, na internet, vemos que as literaturas brasileira e portuguesa levam um atraso enorme (com desvantagem ainda maior para nossos patrícios d'além-mar).

Ora, além da pouca quantidade de obras digitalizadas, há ainda que se chamar a atenção para a baixa confiabilidade dos acervos disponíveis, sério problema para quem se interessa pela confiabilidade do material posto à disposição dos leitores. É óbvio que isso não se restringe ao meio digital. A tradição impressa nunca deixou de interferir na obra escrita pelos literatos. Algumas gralhas tornaram-se até históricas, pelo ridículo a que se expuseram gráficas e editoras. José Mindlin compraz-se em contar o caso de certa edição de Machado de Assis, composta em uma gráfica francesa que trocou um “e” por um “a” na expressão “teve cegada a vista”. O resultado foi um engraçado “teve cagada a vista”. Outro caso digno de nota é o de Fernando Pessoa e de seu heterônimo Alberto Caeiro. Até hoje, são muitas as citações dos versos “Sinto-me nascido a cada momento / Para a eterna novidade do Mundo...”, sem que se atente para a edição crítica de Ivo Castro que, consultados os manuscritos d’*O Guardador de Rebanhos*, concluiu sem hesitação que se deve ler “Sinto-me nascido a cada momento / Para a completa novidade do Mundo...”

Edições com erros de composição fazem as delícias dos bibliófilos. No entanto, a história é diferente no meio digital. A velocidade com que a informação aí se propaga é alucinante. Como resultado, um pequeno erro se multiplica à exaustão e, à custa de ser repetido indefinidamente, pode acabar ganhando foros de verdade. Tomemos novamente os versos de Caeiro. Se procuramos no *Google* por “a eterna novidade do mundo”, vamos encontrar 829 ocorrências; a busca por “a completa novidade do mundo” dará 6 respostas apenas³. Acima, havíamos mencionado o hábito de vários sítios copiarem obras digitalizadas uns dos outros, sem se ocuparem minimamente da correção do que expõem. Com isso, erros e omissões se multiplicam à exaustão.

A título de exemplo, podemos pesquisar no *Google* os sítios que permitem a leitura d’*O Uruguai* em formato digital, prestando atenção à integralidade da obra. Basílio inseriu vários e importantes comentários, em notas aos versos, que não aparecem em edições mais descuidadas em papel. E o mesmo se repete em praticamente todas as edições eletrônicas. O problema se explica pelo fato de elas partirem do arquivo colocado na internet pela Biblioteca Nacional, ainda que nem todas especifiquem aos leitores a fonte de onde veio o arquivo digital. Eles trazem, além dos cinco cantos d’*O Uruguai*, um soneto de Basílio da Gama dedicado ao Conde de Oeiras (irmão do Marquês de Pombal), um de Joaquim Inácio de Seixas Brandão e outro de Alvarenga Peixoto dedicados ao próprio poema. Há também uma citação da *Eneida*

como epígrafe, como consta na edição original. Mas nada das notas! O que muda, de um sítio para outro, é a maneira de o arquivo ser oferecido ao leitor. Uns trazem os arquivos em formato compacto (usando o *winzip*), isto é, sem que possam ser lidos diretamente. Uns colocam o arquivo em formato PDF, o que permite a leitura direta, mas apenas para aqueles que têm o programa *Acrobat Reader* instalado em seu computador; outros ainda exigem que se instale um programa-leitor específico, o que provoca evidentes desconfianças dos leitores mais experimentados, temerosos de instalar arquivos sem procedência conhecida. Uns remetem a arquivos disponíveis em outro endereço da internete, em portal distinto. Há casos bem curiosos.

O sítio do Iglor³ traz a obra para leitura direta e permite salvar o arquivo zipado (vindo da Biblioteca Nacional), mas afirma que seu autor é um certo Basílio de Magalhães. De todos os sítios, apenas um, disponível em members.lycos.co.uk/danieltext/literatura3.htm, traz notas. Mas não são todas as de Basílio (de fato, menos da metade), misturadas às notas de Ivan Teixeira, pertencentes ao volume que este organizou para a Editora da USP. Nem é preciso dizer que este último sítio nem mesmo mostra as fontes de informação, nem das notas, nem do arquivo digital. Em resumo, temos obras digitalizadas com dados incompletos e errôneos, sem um padrão mínimo que uniformize a edição eletrônica e facilite a leitura. O leitor da internete não tem a mínima garantia de quando e onde vai encontrar arquivos confiáveis apresentando a integralidade das obras. Voltando ao sítio do NUPILL, nossa edição digital d'*O Uruguai*, também tirada da que foi digitalizada pela Biblioteca Nacional, também se ressentia do mesmo problema. Todavia, há alguns anos, foi tomada a iniciativa de completar a edição, adicionando as notas. Isso foi feito, mas, os sítios que habitualmente copiam os arquivos, não mantêm política alguma de verificação da qualidade do material digitalizada, nem se preocupa em atualizar os arquivos. Dessa forma, nossa edição, completa, é a única que traz as notas completas de Basílio da Gama.

Indo ao outro extremo desse espectro, podemos talvez afirmar com mais precisão do que se necessita em termos de obras digitalizadas. Precisamos de edições eletrônicas que estejam disponíveis através de programas abertos (ou livres), sem necessidade de leitores específicos, mesmo esses tipo *Acrobat*, em formato PDF já adotado por inúmeros internautas. Eles liberam a utilização com segundas intenções, forçando que os leitores comecem a utilizar o seu programa-editor, que não é livre para utilização, mas que deve ser comprado. Aliás, é preciso definir os tipos de arquivos preferenciais, de acordo com a intenção e a velocidade de acesso aos documentos: para imagens (gif, jpg etc.), para textos editorados (pdf, html, txt, doc, rtf etc.). É claro que isso acarreta conseqüências comerciais sérias e, muitas vezes, os interesses econômicos se têm sobreposto a critérios técnicos e a políticas públicas de inserção digital.

Precisamos também definir alguns padrões ergonômicos que colaborem com a legibilidade, como o estabelecimento de limites máximo e mínimo para as imagens incorporadas à obra digitalizada. Todavia, tais padrões ergonômicos não podem deixar de lado os usuários portadores de necessidades especiais; para os cegos, por exemplo, deve-se acrescentar às obras digitalizadas para leitura um arquivo em áudio.

Tudo isso aponta para o objetivo de dotar o país de boas edições digitais, o que implica obras literárias disponíveis em sua integralidade, dotadas de uma confiabilidade mínima quanto ao conteúdo, o que apenas pode ser feito a partir de uma edição crítica ou de uma outra estabelecida como referência por algum critério de publicação. Foi isso que tentou a Biblioteca Nacional, quando pensou em definir a última edição publicada em vida do autor, como fonte para suas digitalizações. Mas a confiabilidade das edições eletrônicas nunca será efetiva, a menos que se estabeleçam padrões e fontes de confiabilidade para a edição digital de obras literárias. Em algum momento, e que seja o quanto antes!, vamos ter de encarar o desafio de constituir grupos ou organismos encarregados de propor políticas gerais de confiabilidade para edições eletrônicas, a exemplo do que já se faz quanto à construção e à disposição dos metadados para bibliotecas digitais. Por ora, uma medida simples estabeleceria alguma ordem: no caso de sítios que apenas copiam arquivos de outros locais, seria interessante que sempre se informasse a fonte imediata e a fonte primeira da digitalização (já que aquela pode facilmente ter desaparecido). Isso permitiria, se não controlar a qualidade da digitalização, ao menos reconstituir o percurso das eventuais incorreções acrescentadas à obra impressa original.

Todavia, isso tudo que acima se discutiu é, na verdade, apenas o ponto de partida. Queremos mais e precisamos de mais! Queremos ter acesso não apenas a uma edição confiável, mas a várias: primeiras edições ou edições *princeps*, edições tiradas em vida do autor, edições críticas, edições com grafia atualizada contraposta à grafia original... Com isso, teremos a possibilidade de comparar diferentes versões impressas de uma mesma obra, fornecendo um amplo material aos críticos do manuscrito e da ecdótica.

E ainda: precisamos urgentemente que também catálogos de importantes acervos sejam digitalizados, como medida necessária (mas não suficiente) de preservação de patrimônio histórico, artístico e cultural seriamente ameaçado.

De outro lado, é preciso também que tenhamos acesso a ferramentas pedagógicas (glossários técnicos de teoria e de história da literatura, tesouros temáticos etc.) associadas aos conteúdos digitalizados. Elas permitiriam não apenas a leitura das obras digitalizadas, mas sua utilização em estratégias de ensino e de aprendizagem, em ambientes coletivos ou individuais.

Queremos e precisamos de fontes de informações contextuais (históricas, sociológicas, artísticas, bibliográficas etc.) das obras literárias, com acesso rápido e conteúdo confiável. Precisamos de informações acerca das obras e dos autores, como bibliografias críticas e dados de história literária, tudo colocado em ambiente digital. Porém, para isso, ainda teremos de lutar com grandes dificuldades, mesmo antes de chegar ao estágio em que tudo estará colocado na internet. Primeiramente, o acesso a boa parte dessas informações não é imediato, pois elas se encontram em arquivos de acesso mais ou menos restrito, implicam condições precaríssimas de manuseio, trazem marcas às vezes irremediáveis de deterioração. Em segundo lugar, é preciso ainda estar atento a erros constantes. Podemos tomar como exemplo o caso de Olavo Bilac. Quando fomos inserir seus dados bibliográficos em nosso banco de dados, fomos colhê-los na *Obra Reunida de Olavo Bilac*, da Editora Nova Aguilar. Ora, entre outras coisas, deparamos aí com informações incompletas sobre obras em colaboração com outros escritores, havia datas erradas etc. E uma consulta ao banco de dados da Biblioteca Nacional não trouxe ajuda mais substancial, pois também ali as informações eram incompletas.

Tentemos resumir essa aventura que é a digitalização de obras da Literatura Brasileira. Os problemas começam pelas dificuldades de consultar as boas edições, os originais ou os manuscritos. Temos custos de deslocamento tais que, para pesquisadores fora do eixo Rio-São Paulo, quase chegamos a pensar que esse tipo de atividade é inviável. E ainda, chegando às bibliotecas que possuem o acervo a ser consultado, deparamos com proibição de fotocópia e de escanerização, nos vemos às voltas com custos absurdos de microfilmagem ou de fotografia digital. E tudo culmina com falta de pessoal que possa dar uma estrutura mínima de apoio à pesquisa e ao pesquisador. E mesmo quando um projeto de digitalização de acervo literário se realiza em uma de nossas grandes bibliotecas, nem assim o sucesso é garantido. Como dito acima, há alguns anos, a própria *Biblioteca Nacional* deu início à digitalização de obras literárias, apoiada por uma companhia de telecomunicações recém-privatizada, num projeto em parceria com a *PUC-RJ*. Chegou-se a digitalizar 168 obras, mas a coisa parou por aí, sem que a utilização seguida, por parte de leitores, alunos e professores, desse algum retorno ao projeto, permitindo adaptações e mudanças. Algumas dificuldades não foram sanadas: a utilização das páginas do portal da *BN* ainda me pareceu um pouco complexa para o usuário médio; além disso, o fato de o banco de dados e os arquivos com as obras digitalizadas estarem em servidores diferentes (estes, na *PUC-RJ*; aquele, na *BN*), adicionava a essa complexidade uma lentidão desanimadora. Eu disse “adicionava”, pois aparentemente isso foi corrigido, e a *BN* tem feito um esforço louvável na digitalização (mesmo que em formato imagem) de obras.

De toda maneira, o que se vê é uma grande falta de comunicação entre as pessoas e as instituições envolvidas, formando um triângulo com os vértices olhando apenas para fora. De um lado, pesquisadores e universidades, ou seja, quem está em condições ou tem a

obrigação de estabelecer critérios e prioridades para a digitalização de obras e acervos literários; de outro, as bibliotecas com seu acervo e fonte das futuras obras digitalizadas; de outro, órgãos governamentais e agências de fomento, ou seja, a quem compete definir prioridades e induzir, apoiar, financiar áreas prioritárias de pesquisa e desenvolvimento. Talvez aí, nessa distância toda entre eles, esteja a explicação para a falta de um programa consistente de digitalização de obras literárias. Digo isso, em virtude de os custos serem diminutos (recursos humanos, equipamentos, programas, instalações), irrisórios mesmo, diante da importância do projeto e do que se investe em outras áreas. Algumas iniciativas são claramente meritórias. Por exemplo, quanto deve ter custado apenas o prédio destinado a abrigar parte da biblioteca de José Mindlin, na USP? É claro que é um custo mais do que justificado pela extrema importância desse acervo. Mas é um custo muito, muitíssimo mais alto do que se gastaria para colocar 5.000 (isso mesmo, cinco mil) obras na internet. Dessa forma, apenas um esforço coordenado de grandes bibliotecas públicas, pesquisadores da área, centros e núcleos de pesquisa, universidades, órgãos públicos, poderá dar algum resultado. Mas, para chegar a isso, nós pesquisadores devemos perder o medo de democratizar o acesso às informações. Nesse momento, é preciso ir além dos grupos de pesquisa, dos departamentos isolados, dos colóquios e seminários, para unir esforços e planejar com urgência uma política nacional de acervos digitais. A partir dela, então, poderemos propor ações e projetos aos ministérios e órgãos envolvidos, exigindo deles apoio concreto e imediato. Apenas aí, se poderá contar com um acervo de obras digitalizadas capaz de nos ajudar nessa luta terrível contra a falta de leitores e de leituras nesse nosso país.

Notas

¹ Universidade Federal de Santa Catarina. Núcleo de Pesquisas em Informática, Literatura e Linguística. Pesquisador do CNPq.

² É conveniente não esquecer que tais dados se alteram a cada instante. Essas buscas foram todas feitas no dia 05 de dezembro de 2006. Amanhã, provavelmente, esse número será maior.

³ E, se descartamos uma delas, de minha autoria, restam apenas cinco.

³ www.ig.com.br/paginas/novoigler/livros/uraguai_basiliodemagalhaes/index.html