

Avaliação da recuperação da informação em acervos digitais de jornais

Luana Carla de Moura dos Santos

Mestranda; Universidade Federal de Santa Catarina
luana-moura@hotmail.com

Angel Freddy Godoy Viera

Doutor; Universidade Federal de Santa Catarina
recinfor@gmail.com

Resumo: Este estudo avalia o Sistema de Recuperação da Informação de três jornais brasileiros que disponibilizam reproduções fiéis das publicações diárias que fizeram história na sociedade brasileira e internacional. Para a avaliação dos acervos digitais, foi utilizada uma adaptação da metodologia desenvolvida por Sandusky (2002) para bibliotecas digitais. A análise dos resultados permitiu evidenciar que os acervos digitais de jornais necessitam de aperfeiçoamento em relação aos processos que objetivam a recuperação da informação. A definição de metadados precisa contemplar, além da descrição dos conteúdos, relações associativas que considerem as mudanças ortográficas ocorridas durante o período em que os conteúdos dos jornais foram elaborados, devido ao contexto histórico dos documentos. Padrões de interoperabilidade, preservação e iniciativas de acesso aberto também são indispensáveis. A conclusão aponta os desafios dos acervos digitais como importantes fontes de informação, que permitem, além do acesso a documentos primários, a preservação do patrimônio cultural e histórico de uma nação.

Palavras chave: Acervo digital. Jornal. Recuperação da Informação. Digitalização de documentos. Bibliotecas digitais.

1 Introdução

Transitamos em um cenário em que o excesso de informações, ao mesmo tempo em que se torna enriquecedor – devido à disponibilidade de acesso a grandes volumes de informação –, também pode gerar um grande desconforto aos pesquisadores. Essa problemática é evidenciada quando analisados os mecanismos de busca da informação. Muitos não oferecem os requisitos necessários para que a recuperação seja feita com êxito, e informações relevantes sejam alcançadas.

Os Sistemas de Recuperação da Informação (SRI) têm como premissa fornecer mecanismos que dão suporte às buscas realizadas por usuários em determinados ambientes (web, repositórios, base de dados, banco de dados), possibilitando assim que as informações relevantes para os usuários,

disponibilizadas nestes ambientes, sejam encontradas. Choo (2006) acrescenta ainda que a informação recuperada pelo usuário, além de suprir necessidades informacionais, também tem como finalidade “construir significado” e “produzir novo conhecimento”. Assim sendo, se o usuário encontrar a informação desejada, será oportunizada a geração de novo conhecimento e quiçá sua disponibilização na rede.

O avanço tecnológico também viabilizou a disseminação de um grande volume de informações disponibilizadas ao usuário, e a Internet, alavancada pelo rápido desenvolvimento das Tecnologias da Informação e Comunicação (TICs), evidenciou a necessidade de aperfeiçoamento dos Sistemas de Recuperação da Informação.

De acordo com Ferneda (2006, p. 15), uma das funções principais de SRI é apresentar os documentos resultantes de forma que os usuários sejam capazes de verificar facilmente sua pertinência:

Os sistemas de recuperação da informação devem representar o conteúdo dos documentos do *corpus* e apresentá-los ao usuário de maneira que lhe permita uma rápida seleção dos itens que satisfazem total ou parcialmente à sua necessidade de informação, formalizada através de uma expressão de busca.

Outro fator que deve possibilitar uma interação amigável entre sistema e usuário é a interface do SRI. Conforme Pereira (2011, p.31), a interface é o primeiro contato do usuário com a informação desejada. E é por meio dela que ele irá navegar e fazer uso dos SRI:

Em sistemas de busca baseados em computador, a interface é o ponto de partida para o processo de busca e recuperação da informação, fundamental para garantir a excelência do diálogo entre o sistema e o humano. A interface é, portanto, a responsável por fazer a mediação entre a visualização e a representação de redes de informações, permitindo, dessa maneira, que o usuário siga as ligações entre os conceitos que estão sendo explorados.

Entretanto, por mais que os SRI se aperfeiçoem e propiciem sistemas que demandam menos do melhores resultados letramento digital, ainda são

dependentes e necessitam da contribuição dos usuários para obterem na Recuperação da Informação (RI). Segundo Lopes (2002, p. 65), as dificuldades do usuário inerentes ao SRI ainda não foram solucionadas pelas tecnologias:

Apesar dos intensivos programas de treinamento oferecidos pelos produtores das bases de dados, pelos próprios sistemas de recuperação em linha, de toda a documentação existente sobre as características de cada base de dados e suas respectivas estruturas de informação, dos sistemas amigáveis que oferecem "menus" para guiar o usuário em cada etapa do processo de busca, das linguagens de busca com recursos especiais para se aproximarem cada vez mais do usuário inexperiente, o processo de busca continua sendo um fator de dificuldade que ainda não foi minimizado pelas novas tecnologias disponíveis.

Neste âmbito, estudos que avaliem estas interações e as constantes dificuldades apresentadas pelos usuários devem ser explorados, a fim de elevar os resultados relevantes nas buscas. Na visão de Saracevic (1996, p.45), os estudos sobre SRI devem ter como foco os seguintes aspectos:

[...] o comportamento humano frente à informação; a interação homem-computador, com ênfase no lado humano da equação; relevância, utilidade, obsolescência e outros atributos do uso da informação juntamente com medidas e métodos de avaliação dos sistemas de recuperação da informação; economia, impacto e valor da informação, dentre outros.

Portanto, propõe-se com este artigo, avaliar, na visão do usuário da informação, o SRI de três jornais brasileiros que disponibilizam em seus bancos de dados reproduções digitais dos jornais que fizeram história na sociedade brasileira. A avaliação será feita por meio de uma adaptação da metodologia proposta por Sandusky (2002) para bibliotecas digitais. Para que seja feita essa adaptação, primeiramente, serão conceituadas e apresentadas as semelhanças e diferenças de bibliotecas digitais e acervos digitais. Em um segundo momento, será analisada e diferenciada a dimensão de jornais impressos digitais e jornais *on-line*. E, por último, serão apresentados os recursos oferecidos pelos acervos digitais disponibilizados pelos jornais pesquisados neste estudo.

2 Bibliotecas Digitais e Acervos Digitais

O conceito de bibliotecas digitais, inicialmente, caracteriza-se como um ambiente onde são armazenados documentos primários, que ficam disponíveis para acesso em formato digital. No entanto, desde o seu surgimento, as bibliotecas digitais estão em constante evolução tecnológica e reúnem “[...] acervos, serviços e pessoas para estabelecer um ciclo de vida completo de criação, disseminação, uso e preservação de dados, informações e conhecimento.” (INTERNATIONAL FEDERATION..., p. 2, 2012).

Deste modo, é possível também destacar a biblioteca digital em dois sentidos. Primeiramente, sob a visão dos profissionais da área da Ciência da Informação. Segundo Sayão (2009, p. 9-10), esses profissionais “[...] enxergam a biblioteca como uma extensão dos sistemas de computadores em rede”, mas o autor também destaca que a biblioteca pode apresentar a incumbência de disponibilizar novas publicações na web:

Para o bibliotecário "a biblioteca digital é um estágio a mais no desenvolvimento contínuo de novos meios de publicação, em que a biblioteca soma a responsabilidade de também ser uma publicadora web, bem como uma nova infra-estrutura tecnológica e organizacional voltada para potencializar a sua missão de disseminar informação e conhecimento. (SAYÃO, 2009, p. 9-10).

A biblioteca digital também pode ser analisada sob as perspectivas dos profissionais da área de Ciência da Computação. Estes, de acordo com Duguid (19971 apud Sayão 2009, p. 10) enxergam as bibliotecas digitais como um ambiente de variadas coleções, sendo um “Ambiente distribuído que integra coleções, serviços e pessoas na sustentação do ciclo de vida completo de criação, disseminação, uso e preservação de dados, informação e conhecimento”.

O acesso universal, sem nenhum tipo de restrição, é a direção que devem seguir as bibliotecas digitais (BORGES, 2003, p. 655). Esses ambientes precisam viabilizar a troca de conhecimento, e a forma como esta interação acontece não deve ser circunvalada de limitações. Conforme Sayão (2009, p. 7), as bibliotecas

digitais possuem compromisso com “[...] a disponibilidade crescente de conteúdos digitais em escala planetária [e] a possibilidade de digitalização a um custo economicamente viável de conteúdos em mídias convencionais [...]”, assim como com a facilitação de concepção de novos serviços, de forma a integrá-los nos sistemas digitais.

Contudo, o que se pode evidenciar é que muitas coleções, associadas aos distintos conteúdos, formatos e apropriações, são disseminadas fora das bibliotecas e controladas por instituições privadas, públicas e de diversos segmentos. De acordo com Arellano (2004, p. 15), com o crescimento da produção digital, os questionamentos sobre sua disponibilização e preservação são mais constantes:

Atualmente muitas coleções digitais importantes estão sendo construídas fora das bibliotecas por diferentes organizações, ou sendo publicadas diretamente na Internet. Com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de se ter garantida a sua disponibilização e preservação por longos períodos de tempo.

Houve avanços em diferentes sentidos na conceituação de bibliotecas digitais, e a ideia de que ela deve abranger e ampliar diferentes serviços é predominante. Quando relacionamos os serviços de uma biblioteca digital com os oferecidos por um acervo digital, os termos são tratados com similaridades.

Abordagens que promovem a biblioteca digital em um sentido mais amplo, oferecendo diferentes serviços, e os acervos digitais, direcionados a documentos digitalizados e disponibilizados em ambientes digitais, são frequentemente adotadas.

A Fundação Biblioteca Nacional pode ser citada como exemplo das semelhanças existentes entre biblioteca digital e acervo digital. A instituição oferece uma plataforma denominada de Biblioteca Nacional Digital Brasil, na qual comporta seções denominadas de: Artigos, Dossiês, Exposições e Acervo digital.

Nas seções é possível localizar coleções de documentos digitalizados, transcritos, relatos de exposições, digitalizações de obras artísticas, dossiês de

autores ilustres e personalidades e notícias de mídias jornalísticas referentes aos autores e personalidades.

Na seção referente ao acervo digital, é oferecido um SRI em que é possível unir diferentes métodos de busca (simples e avançada) para localizar documentos nos diferentes tipos de acervos. São eles: gravura, icnográfico, cartográfico, livro, manuscrito, partitura, periódico, jornal, sonoro e vídeo. Dentro de cada acervo, ainda é possível selecionar o tipo de documento: fotografia, folheto, livro, desenho, entre outros.

Assim, além de disponibilizar os documentos em formato digital, a Biblioteca Nacional Digital também tem a preocupação de propiciar a organização dessas informações dentro do acervo digital, assim como possibilitar a recuperação das informações contidas nesses documentos.

Essa preocupação está presente também na Carta de Apresentação da Biblioteca Nacional Digital, disponível em sua página. Nela, são mencionadas as mudanças sucedidas pela inovação digital, que impactaram de modo direto na recuperação da informação, conforme a citação abaixo:

O mercado pode transferir para a inovação do digital os velhos dogmas do funcionalismo, que as ciências sociais tomaram de empréstimo às ciências naturais. E tende a deixar esquecida a diversidade das línguas naturais em favor de um algoritmo único, responsável por todos os mecanismos de busca. Este tipo de preocupação está presente em nossa arrancada rumo à digitalização do acervo. A democratização do acesso à memória cultural não deve furtar-se à pesquisa de uma orientação nacional para os mecanismos de busca. (BIBLIOTECA NACIONAL..., 2014, doc. eletrônico).

A aplicação das linguagens naturais e o desenvolvimento de mecanismos que sejam capazes de atender a diferentes contextos demonstrados nas necessidades informacionais dos usuários, são destacados pela Biblioteca Nacional Digital como recursos indispensáveis em acervos digitalizados, uma vez que facultam que esses ambientes sejam mais procurados pelos usuários, pois facilitam a recuperação da informação.

No programa Sociedade da Informação no Brasil - Livro verde, iniciativa do Ministério da Ciência e Tecnologia, foi elaborado um capítulo sobre o tema Conteúdos de Identidades (TAKAHASKI, 2000), em que algumas recomendações

são sugeridas em referência aos conteúdos digitais. Entre elas, destaca-se a avaliação de processos que visem o armazenamento, recuperação da informação e difusão de conteúdos de valor histórico, conforme abaixo:

- a) propor o estabelecimento de normas técnicas para o tratamento de conteúdos (metadados), para garantir maior racionalidade nos processos de armazenamento e maior pertinência e relevância na recuperação de informações;
- b) desenvolver uma legislação adequada de linhas de fomento para a digitalização e disponibilização de conteúdos de interesse cultural, a exemplo da Lei Rouanet (BRASIL, 1991);
- c) instituir um programa de digitalização de obras de valor histórico;
- d) patrocinar o inventário e a digitalização de obras literárias, manuscritos, gravuras e todo material iconográfico sobre o Brasil dos diferentes períodos históricos, como meio de resgate da memória da formação nacional;
- e) fomentar a tradução (e as tecnologias de processamento de linguagem natural que podem apoiar a atividade) de conteúdos entre a língua portuguesa e outras línguas na Internet.

Expandir a ideia de que bibliotecas e acervos digitais devem propiciar o intercâmbio de informações amplifica a concepção de que, nesses ambientes, o que prevalece é o acesso ao conhecimento. Em bibliotecas ou acervos, a avaliação do ambiente deve levar em consideração critérios similares. Entre eles, a recuperação da informação, usabilidade, audiência, acesso e serviços.

3 Jornal Impresso Digital Versus Jornal *On-line*

De acordo com informações retiradas da cronologia da Associação Nacional dos Jornais (ANJ), a história do jornal no mundo inicia-se em 59 a.C. (ASSOCIAÇÃO NACIONAL..., c2008). A sua invenção é atribuída ao Imperador Romano Júlio César. Na ocasião, o Imperador ordenou que os

principais acontecimentos políticos do dia fossem publicados e divulgados para seu império em um documento chamado de Acta Diurna.

As transformações do papel impresso passaram pela prensa de Gutemberg, pela invenção do telégrafo, pelo surgimento do rádio e televisão e, nas últimas décadas, pela popularização da Internet.

Acreditava-se que, com a chegada das TICs, o jornal se tornaria obsoleto. No entanto, a mídia impressa se adaptou, conciliando suas tradicionais características às novas ofertas da rede e, possibilitando assim, um acesso facilitado e uma audiência mais veloz.

Deixando de ser impresso e tornando-se digital, o jornal se configura em três fases no ambiente virtual. Primeiro, temos a opção de digitalização, que seria uma reprodução da versão impressa para o digital. É possível também atribuir a essa primeira dimensão o formato *Flip*, que se apresenta de forma equivalente à versão impressa em formato digital. O Flip permite ao usuário visualizar o conteúdo do jornal de forma semelhante ao convencional – as páginas são viradas a partir de um clique, e a estrutura e organização das informações são apresentadas conforme o que foi disponibilizado no formato em papel. Nesta perspectiva, o leitor pode optar por receber diariamente a versão impressa ou fazer a assinatura digital, e assim consumir o conteúdo por meio de um computador –, enfatizando que os conteúdos serão os mesmos, alterando apenas o suporte no qual serão disponibilizados.

Nesse primeiro cenário ainda é possível identificar uma subdivisão: os jornais que nasceram em papel e posteriormente foram introduzidos ao ambiente digital, e os jornais que já nasceram digitais e também foram disponibilizados em formato impresso.

Na versão digital, os jornais ganham inúmeras vantagens, se comparados ao convencional. Os custos de produção e distribuição são reduzidos e o acesso é ampliado, uma vez que não existe fator limitador – exceto o fator tecnológico – em relação ao número de leitores que acessam uma mesma publicação. A conservação e preservação são facilitadas pelos bancos de dados dos próprios jornais; e a busca é simplificada, com sistemas eficazes de recuperação da informação que retornam em tempo ágil a informação desejada pelos leitores.

Em um segundo momento, temos o jornalismo que ainda tem transposições do impresso, mas já introduz alternativas tecnológicas oferecidas na web. A possibilidade de edição e publicação de novas notícias, instantaneamente, sem precisar esperar pela edição do dia seguinte, foi uma grande conquista do jornalismo na web. Este cenário é abordado por Palacios et al. (2002, p. 3):

Com o aperfeiçoamento e desenvolvimento da estrutura técnica da Internet, pode-se identificar uma segunda fase, a da metáfora, quando, mesmo 'atrelado' ao modelo do jornal impresso, os produtos começam a apresentar experiências na tentativa de explorar as características oferecidas pela rede.

A terceira fase acontece somente e diretamente na web. O jornalismo construído *on-line*, muitas vezes com a cooperação do usuário, de acordo com Palacios et al. (2002), apresenta cinco características: multimídia/convergência, interatividade, hipertextualidade, personalização e memória.

Diante das possibilidades ofertadas pelo jornal no ambiente digital, o presente artigo analisa acervos digitais que têm sua concentração voltada para jornais impressos em meio digital, englobando, neste sentido, jornais que foram digitalizados e submetidos às plataformas digitais, e também jornais que nasceram digitalmente, mas possuem as características tradicionais de uma única publicação diária.

4 Recursos de Recuperação da Informação em Acervos de Jornais

As possibilidades de utilização de acervos digitais de jornais são diversificadas. Os ambientes virtuais que servem de acesso à informação e também de preservação do patrimônio cultural e informacional estão em constante desenvolvimento. No Brasil, por volta do ano de 2012, os jornais com maior circulação disponibilizaram na Internet seus acervos, compostos por edições publicadas ao longo da existência das mídias jornalísticas.

Os recursos de RI mais frequentes disponibilizados nos acervos dos jornais são:

- a) **busca rápida** - possibilita a busca por palavras, seções do jornal (esportivo, econômico, policial) e períodos de tempo (dia, mês, ano);
- b) **busca detalhada** - possibilita a relação de palavras chaves, a exclusão de termos e a busca por mais de uma seção, além da busca por períodos aproximados ou datas exatas;
- c) **página de resultados** - os resultados das buscas exibem quantas ocorrências foram recuperadas como resposta a uma consulta do usuário, assim como permitem a aplicação de filtros, refinando a busca por período, ano, mês e dia, além da análise por métricas;
- d) **navegação intuitiva** - possibilita que o usuário faça a escolha de qualquer data disponível no acervo para a pesquisa;
- e) **busca da informação por tópicos** - busca de termos catalogados que unem fotos, vídeos e páginas do jornal acerca de um mesmo tema. As notícias nesta função são retiradas das páginas dos jornais e transformadas em textos semelhantes ao de uma enciclopédia;
- f) **conversor de valores** - convertem os valores e custos impostos em serviços, bens e consumos destacados nas páginas dos jornais em outras décadas para os dias atuais;
- g) **expansão automatizada da consulta dos termos em português da ortografia vigente para ortografias antigas** - possibilita que termos atualizados ou modificados na língua portuguesa sejam recuperados de acordo com sua gramática atual. Ex.: “Pharmacia” e “Farmácia”;
- h) **armazenamento de documentos** - possibilita, além da impressão do documento, armazenamento em nuvem ou em computador local;
- i) **zooming** - permite ao usuário ampliar a página selecionada e visualizar as informações com mais detalhamento;
- j) **busca de palavras dentro do texto** - possibilita a localização de palavras dentro do texto do jornal que está sendo visualizado.

Esses recursos devem favorecer a interação do usuário e facilitar a busca e o acesso à informação pesquisada no acervo do jornal. Portanto, a análise e a atualização dos mecanismos de busca, conforme as necessidades evidenciadas pelos

usuários e pelo contexto a ser explorado, devem ser constantemente revisadas e adaptadas.

5 Procedimentos Metodológicos

Esta pesquisa caracteriza-se como exploratória e descritiva, com abordagem qualitativa. A coleta dos dados aconteceu em julho de 2014, com a finalidade de analisar os recursos de pesquisa de três jornais brasileiros: A Folha de S. Paulo, fundado em 1921, O Globo, fundado em 1925, e O Estado de S. Paulo, fundado em 1875. A escolha destes jornais justifica-se pela longa data de existência que possuem no cenário jornalístico brasileiro, pela expressividade que apresentam perante a sociedade brasileira e mundial, por deterem os maiores números de circulação no país, de acordo com a ANJ (ASSOCIAÇÃO NACIONAL..., c2008), e também por disponibilizarem seus acervos digitais na Internet, possibilitando assim esta pesquisa.

O objetivo da análise é apontar as possibilidades de busca dos acervos digitais e apresentar recomendações de melhoria, no que se refere à recuperação da informação.

As análises contemplam versões de jornais impressos que foram digitalizados e submetidos ao ambiente digital, e também jornais que possuem as características tradicionais dos jornais impressos, mas já nasceram digitais e estão disponíveis para pesquisas nos acervos dos jornais que fazem parte da amostra.

Para avaliação do acervo, foi utilizada a metodologia desenvolvida por Sandusky (2002) para bibliotecas digitais. Entretanto, a metodologia foi adaptada para que possa atender aos critérios propagados pelos acervos digitais. Por esta razão, os critérios avaliados são:

- a) **direitos de acesso** - Quem pode acessar o conteúdo? Condições de uso e suporte de acesso ao conteúdo. Acesso grátis ou necessita de assinatura? Pagamento direta ou indiretamente? Visibilidade do conteúdo completo ou acesso condicionado?;

- b) **conteúdo** - Qual a extensão dos conteúdos? O conteúdo é digitalizado a partir de material não-digital existente ou o conteúdo é criado com informação digital? O conteúdo é de domínio público ou restrito por direitos autorais? O conteúdo é criado e controlado por única entidade e/ou múltiplas fontes?;
- c) **recursos de recuperação da informação** - Oferece diferentes tipos de recursos de buscas e/ou consultas (simples, detalhada)?;
- d) **serviços** - Suporte para interação usuário-sistema? Serviço de referência para contato com especialistas em tempo real? Quais são os serviços de seleção e organização de conteúdo? Quais são as políticas de indexação?;
- e) **design e desenvolvimento** - (processo de construção e manutenção) Acervo desenhado tendo em mente os outros critérios avaliados: acesso, conteúdo, além da audiência (público-alvo). *Design interativo* utiliza os *feedbacks* dos usuários?

Por fim, o tratamento dos dados levantados será exposto em forma de quadros, devidamente estruturados, conforme os critérios analisados.

6 Resultados e Discussão

Neste tópico serão levantadas as funcionalidades avaliadas nos acervos digitais dos jornais, Folha de S. Paulo, O Estado de S. Paulo e o Globo, analisadas conforme a adaptação dos critérios da metodologia proposta por Sandusky (2002) – direitos de acesso, conteúdo, recursos de recuperação da informação, serviços e design, e desenvolvimento.

6.1 Folha de S. Paulo

A coleção de jornais da Folha de S. Paulo disponibilizada na internet contempla 93 anos de história. A navegação no acervo permite selecionar a pesquisa em diferentes coleções: Folha de S. Paulo, Folha da Manhã e a Folha da Noite, que

seriam edições distintas distribuídas diariamente ao público. A primeira edição disponível no acervo é datada com o ano de 1921. (FOLHA DE SÃO PAULO, [200-]).

Aplicando os critérios adaptados da avaliação estabelecidos por Sandusky (2002) ao acervo digital do jornal da Folha de S. Paulo, temos as seguintes análises:

Quadro 1 - Avaliação SRI acervo Folha de S. Paulo

Critério	Avaliação
Direitos de acesso	Qualquer pessoa pode acessar o conteúdo. Contudo, o acesso é condicionado. É disponibilizado um período de degustação, em que o usuário pode fazer pesquisa e acessar o conteúdo sem assinatura. Após esse período, é solicitada a assinatura, que pode ser feita diretamente no portal. O serviço ofertado não é exclusivo para o acesso ao acervo. Para adquirir a permissão de acesso, é necessário fazer a assinatura da versão impressa e/ou digital. A assinatura permitirá a liberação de acesso às antigas publicações.
Conteúdo	As páginas dos jornais disponíveis para pesquisa são digitalizadas de material não digital. Também estão disponíveis edições que já nasceram digitais. Os conteúdos são disponibilizados e mantidos pela entidade. Desta maneira, possuem direitos reservados, ficando expressamente proibida a reprodução em qualquer meio de comunicação (eletrônico ou impresso) sem autorização da Folha de S. Paulo.
Recursos de Recuperação da Informação	Possui a busca simples, que permite recuperar a edição por ano, mês ou dia de publicação. Também possibilita a busca detalhada, que permite a pesquisa por palavras-chave, frases, ou a exclusão de palavras e frases específicas na pesquisa, além de oferecer a busca por períodos ou datas exatas e também por temas (ex.: política, economia, cultura e sociedade, entre outros).
Serviços	Permite somente acesso ao conteúdo. Não possui contato instantâneo com os especialistas. É ofertado um tutorial que explora como podem ser usados os serviços oferecidos pelo acervo. Dentre os serviços de armazenamento estão: <i>download</i> do documento, envio por e-mail, compartilhamento nas redes sociais e opção de salvar o arquivo em computador local ou armazenar em nuvem. A organização dos documentos é ordenada pelos jornais (Folha de S. Paulo, Folha da Manhã e Folha da Noite), ano, mês, dia, seção (Esporte; Economia e Finanças; Mulher e Lar, entre outros) e página do jornal.

Design	<p>O acervo contempla algumas coleções organizadas para atrair o usuário. Entretanto, o acervo foi criado tendo em mente as funcionalidades de um jornal tradicional, fazendo uso das contribuições das tecnologias e tendo como prioridade a busca e recuperação da informação. O acesso ao conteúdo, por meio das buscas simples e detalhada, é posicionado em destaque, possibilitando que estes serviços sejam valorizados dentre as funções do acervo. As interfaces de interação com o utilizador permitem conhecer o que existe na base de dados documental e também apresentam como se deve realizar a consulta e analisar os resultados obtidos. As consultas facilitam a memorização, e a disposição do conteúdo é acessível. Na interface inicial, os recursos para a busca simples no acervo (jornal, data e seção) e a busca detalhada estão bem estruturados e permitem que o usuário defina todos os seus interesses de pesquisa já no primeiro acesso. Na interface que traz o documento recuperado na busca, é possível visualizar uma única página de uma edição do jornal, mas é possível passar para a página seguinte, assim como modificar a data e o jornal selecionado. Os serviços de compartilhamento, impressão, informações sobre o acervo e contato com os responsáveis pelo acervo estão em destaque, facilitando a navegação do usuário.</p>
---------------	---

Fonte: Elaborado pelos autores

A dificuldade mais significativa encontrada na avaliação dos recursos de pesquisa do acervo foi a ausência da possibilidade de localização de palavras dentro do texto que está sendo visualizado, devido à parte das notícias digitalizadas não serem pesquisáveis.

6.2 O Estado de S. Paulo

A coleção do acervo digital do jornal O Estado de S. Paulo, denominado também de Acervo Estadão, engloba um arquivo *on-line* de todas as edições publicadas desde 1875, somando 139 anos de história (ESTADÃO, c2007).

Na pesquisa do acervo, é possível navegar em reportagens e fotos publicadas, além de ter disponíveis os períodos de publicação de jornais que foram censurados durante a ditadura militar. Podemos comparar a versão censurada com a que foi publicada e distribuída à sociedade, permitindo assim que as memórias dos fatos acontecidos durante esse período da história do Brasil sejam guardadas.

As edições mais recentes entram para o acervo após 30 dias da publicação. Antes disso, elas ficam disponíveis apenas para assinantes.

Aplicando os critérios adaptados da metodologia proposta por Sandusky (2002) no acervo digital do jornal O Estado de S. Paulo, obteve-se as seguintes análises:

Quadro 2 - Avaliação SRI acervo O Estado de S. Paulo

Critério	Avaliação
Direitos de acesso	Qualquer pessoa pode pesquisar o conteúdo do acervo. Contudo, o acesso completo dos documentos é restrito para assinantes dos jornais impresso e digital.
Conteúdo	As páginas dos jornais disponíveis para pesquisa são digitalizadas de material não digital. Também estão disponíveis edições que já nasceram digitais. O material é mantido pelo grupo O Estado, vinculado ao jornal O Estado de S. Paulo, e os direitos são reservados.
Recursos de Recuperação da Informação	Possibilita a busca simples por termos, contemplando todo o acervo, ou buscas específicas que trazem somente as capas dos jornais ou edições de materiais censurados. Não oferece a busca detalhada ou avançada. Possibilita também a busca em datas exatas, filtro por tópicos, personalidades e lugares. Permite a busca de palavras dentro do texto do jornal que esta sendo visualizado pelo usuário.
Serviços	Permite somente acesso ao conteúdo. Não possui contato instantâneo com especialistas. O tutorial que descreve como podem ser usados os serviços oferecidos pelo acervo não explica, passo a passo, os recursos de busca e recuperação da informação. Permite a impressão do documento e o <i>download</i> do arquivo em um computador local. Porém, este serviço só está disponível para assinantes. Na busca por tópicos, os textos estão organizados em grupos temáticos que já foram previamente indexados e catalogados. São mais de 400 verbetes catalogados que unem fotos, vídeos e páginas do jornal acerca de um mesmo tema. A organização dos documentos está ordenada por década, ano, mês e dia. Oferece também um serviço de conversor de valores – índice para comparar preços antigos com atuais.

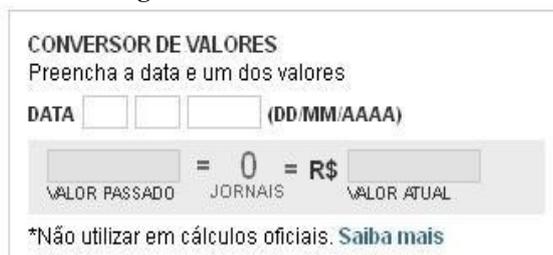
<p>Design</p>	<p>A interface inicial destaca o serviço de busca do acervo, mas também oferece conteúdos distribuídos em coleções bem estruturadas. A interface que traz o documento recuperado possibilita navegar na página do dia do jornal. A visualização é feita página por página. É possível ir para página seguinte, mas só é possível visualizar uma data seguinte realizando uma nova pesquisa. Os recursos de <i>zooming</i>, impressão, <i>download</i> do documento e compartilhamento em mídias sociais está em destaque na página. Os serviços de compartilhamento, impressão, informações sobre o acervo e contato com os responsáveis pelo acervo também ganham destaque, facilitando a navegação do usuário. O acervo não oferece um histórico das buscas realizadas, para que o usuário possa revisar determinado documento, sem a necessidade de fazer uma nova busca.</p>
----------------------	--

Fonte: Elaborado pelos autores

Uma das dificuldades apresentada no Acervo do Estadão é a baixa definição da digitalização, que, mesmo com o serviço de *zooming*, prejudica a visualização clara de alguns textos e imagens das páginas dos jornais devido à pixelização. Outro fator de entrave é a localização de palavras dentro do texto. Mesmo que o acervo ofereça este recurso, a busca de palavras dentro do texto ainda não é satisfatória. Muitas palavras em destaque no texto e visíveis para o usuário, quando são submetidas à busca automática no texto, não são retornadas como termos solicitados.

Um recurso interessante do acervo é o conversor de valores (Figura 1 e 2) – índice criado para comparar preços antigos que estampavam nas páginas dos jornais com os da atualidade. Este serviço é baseado em uma fórmula que inclui dividir o preço do produto publicado pelo preço do jornal na data da publicação (ESTADÃO, c2007). Dessa fórmula, é gerado o resultado que permitirá a atualização do preço: o número de exemplares de jornal que era possível comprar com o valor de determinado produto, multiplicado pelo valor atual do jornal.

Figura 1- Conversor de valores



Fonte: Instruções do acervo O Estado de S. Paulo (ESTADÃO, c2007).

Figura 2- Conversor de valores (cálculo)



Fonte: Instruções do acervo O Estado de S. Paulo (ESTADÃO, c2007).

O jornal O Estado de S. Paulo contempla o mais antigo acervo e oferece também recursos que podem ser utilizados concomitantemente para enriquecer as informações recuperadas no acervo, a exemplo do recurso Conversor de valores.

6.3 O Globo

O acervo do jornal O Globo permite o acesso à versão digital de todas as páginas e matérias publicadas desde sua primeira edição, em 29 de julho de 1925, contemplando 89 anos de história (O GLOBO, 2011). No acervo, além da reprodução fiel dos jornais impressos publicados e distribuídos à sociedade, também é possível encontrar fragmentos das páginas digitalizadas. Os textos foram extraídos com o objetivo de facilitar a leitura e também oferecer coleções de assuntos personalizados aos leitores e usuários do acervo.

Aplicando os critérios de avaliação no acervo digital do jornal O Globo, temos as seguintes análises:

Quadro 3- Avaliação SRI acervo jornal O Globo

Critério	Avaliação
Direitos de acesso	O acesso ao conteúdo é condicionado. É possível realizar as buscas e encontrar o material. No entanto, o serviço de zooming – que permite ampliar a notícia e assim visualizar com clareza o conteúdo – é oferecido somente após o <i>login</i> , mediante assinatura. O acesso completo ao acervo é disponibilizado somente aos assinantes do jornal.
Conteúdo	As páginas dos jornais disponíveis para pesquisa são digitalizadas de material não digital. Também estão disponíveis edições que já nasceram digitais. O material disponibilizado é mantido pelo Infoglobo Comunicação e Participações, e os direitos são reservados.
Recursos da Recuperação da Informação	Possibilita busca por palavras-chave e busca por década, ano, mês e dia. Na busca por palavras-chave, quando obtido o resultado, é possível refinar a busca, unindo outras palavras, excluindo palavras e/ou buscando por expressões ou frases exatas.
Serviços	Permite somente acesso ao conteúdo. O serviço de referência não é em tempo real e a solução disponível é de contato via FAQ, que oferece perguntas e respostas sobre os serviços do acervo e também o envio de perguntas, caso a dúvida permaneça. Possibilita a impressão e o compartilhamento em redes sociais dos conteúdos recuperados. A organização dos documentos é ordenada por décadas, ano, mês e dia, além de por coleções de fatos históricos devidamente indexados, conforme assuntos já determinados.
Design	A interface de interação com o utilizador tem como pontos principais os mecanismos de buscas. A navegação é facilitada pelos links, e a disponibilização dos conteúdos é adequada para uma busca simplificada pelo usuário. A página que apresenta os resultados da pesquisa oferece um serviço ágil de refinamento da busca, que possibilita também analisar os resultados por meio de métricas, que dividem os itens recuperados dentre as décadas do acervo. A interface que traz o documento recuperado permite visualizar página por página o documento e disponibiliza o serviço de <i>zooming</i> . Para que o usuário faça outras buscas, é necessário retornar para a interface anterior aos resultados, dificultando a busca e acarretando um maior número de cliques para chegar a informação desejada.

Fonte: Elaborado pelos autores

Os pontos fortes do acervo são acentuados pela ótima digitalização das páginas e também pela apresentação dos resultados em métricas, que possibilitam que os itens recuperados sejam divididos em décadas (Figura 3). O oferecimento

de atrativos exclusivos para assinantes – como ver a capa do jornal na data de nascimento do usuário (Figuras 4 e 5), frases de personalidades e principais fatos em caricaturas e charges que marcaram o jornal desde 1925 – também trazem um diferencial ao acervo.

Figura 3 - Métricas décadas



Fonte: Acervo do jornal O Globo (2011).

Figura 4 - Capa Jornal o Globo de 19/08/1950



Fonte: Acervo do jornal O Globo (2011).

Figura 5 - Recurso “Veja a capa do jornal no dia em que você nasceu”

VEJA A CAPA DO GLOBO DO DIA EM QUE VOCÊ NASCEU

ANO ▼ MÊS ▼ DIA ▼

Fonte: Acervo do jornal O Globo (2011).

Em relação às dificuldades avistadas no acervo, a ausência da opção de localizar palavras dentro do texto certamente compromete a recuperação das informações. Outro entrave evidenciado é observado na busca por palavras. Os resultados são obtidos, porém não é destacado no texto em que posição a palavra recuperada se encontra.

7 Recomendações

Para que sejam aperfeiçoadas as contribuições dos acervos digitais de jornais à sociedade, algumas melhorias são necessárias. Essas melhorias estão relacionadas, principalmente, à busca e recuperação da informação, que facilitam a navegação e a utilização dos recursos oferecidos pelos acervos digitais.

Entre os principais serviços, destacamos a busca de palavras dentro dos textos – recurso oferecido por apenas um dos acervos avaliados, mas que mesmo assim não apresentava uma recuperação 100% satisfatória. A disponibilização adequada desse recurso irá facilitar o encontro de informações relevantes ao usuário dentro do documento que está sendo visualizado.

Outra sugestão que deve ser avaliada é a padronização dos serviços. Os acervos apresentam o mesmo tipo de material, porém os conteúdos são diferentes, pois se trata de veículos de comunicação distintos. Seria interessante que todos compartilhassem de recursos semelhantes e com a disposição dos conteúdos feita de forma equivalente, conforme bibliotecas digitais, bases de dados e repositórios. A busca simples, detalhada ou avançada e o refinamento da busca são recursos essenciais nos acervos. Portanto, a padronização e oferecimento destes serviços irão facilitar o entendimento do usuário e estimular a navegação em acervos de diferentes veículos de comunicação, diminuindo consideravelmente a curva de aprendizado dos usuários.

É possível também destacar a personalização dos serviços aos usuários, recurso pouco explorado nos acervos. Esta função iria simplificar os processos de busca e, ao mesmo tempo, auxiliar o usuário a encontrar o material desejado. A aplicação de agentes inteligentes no processo de busca de cada usuário, facilitado pelo *login* do assinante, também propiciaria maior compreensibilidade quanto aos

resultados obtidos na busca. Ofereceria, assim, conjuntos de soluções a serem aplicados a cada usuário, conforme o histórico de suas pesquisas, e a permissão da customização da interface para atender o perfil de interesse e de uso dos usuários.

Um fator pouco explorado e que precisa ser melhor desenvolvido é a aplicação da expansão automatizada de consulta dos termos em português da ortografia vigente para ortografias antigas. Um dos acervos chega a ter mais de um século de existência. Neste período, mudanças ortográficas e expressões que derivam de uma mesma definição foram alteradas. Por esta razão, a relação desses termos deve ser garantida pelo SRI, mapeando os termos utilizados na consulta para aumentar a revocação do sistema, de forma que a recuperação das informações não dependa do letramento do usuário.

A adoção de metadados de descrição e de preservação também é indispensável para acervos digitais. Estes padrões contribuem para a interoperabilidade, a recuperação e a organização das informações, além de propiciar que as informações sejam acessíveis a um longo prazo.

8 Considerações Finais

Com base nos quadros dos resultados e atendendo à finalidade do processo de avaliação, conclui-se que os acervos utilizados no estudo se caracterizam como ambientes digitais que permitem apenas o acesso ao conteúdo dos documentos – não possibilitando interações instantâneas com especialistas para questões relacionadas ao próprio conteúdo, ou ao desempenho do sistema, também não viabilizando a comunicação entre usuários, função já difundida no jornalismo online, com a liberação de comentários nas matérias publicadas, o que muitas vezes estabelece um diálogo entre os leitores.

Quanto à utilidade, além do aspecto informativo, verifica-se a importância cultural dos acervos, que além de propiciarem a busca a informações, trazem consigo elementos cruciais para o patrimônio histórico da sociedade. De um modo geral, iniciativas como estas também fomentam projetos que almejam que os conteúdos sejam preservados e adotem a iniciativa do acesso aberto, para que alcancem um maior número de usuários.

A pesquisa na área de acervos digitais, pelo seu domínio amplo e pela sua

complexidade, deve ser explorada. Para tanto, é importante a análise de outros acervos (brasileiros e de outros países), para que, assim, os serviços sejam aperfeiçoados e possibilitem maior acesso às informações, além de identificar padrões que estão sendo utilizados em nível internacional, para viabilizar a interoperabilidade entre sistemas – o que potencializaria o intercâmbio e o acesso por um maior número de usuários.

Estudos relacionados com a preservação e a conservação dos documentos jornalísticos em formato impresso ou digital também necessitam de uma maior discussão. Como assegurar que os documentos que foram guardados por tanto tempo fisicamente, quando introduzidos em ambiente digital, não serão dispersos? A apreensão também é grande em relação aos documentos que foram originados digitalmente, em relação a como preservar o acesso a esses documentos em longo prazo, considerando a rápida mudança das plataformas de hardware, formatos e padrões no ambiente digital. Muitas iniciativas já são desenvolvidas neste sentido, mas no caso dos acervos digitais, de acordo com Dodebei (2006), tanto os documentos nascidos digitalmente quanto os que foram digitalizados, é fundamental que o patrimônio capturado não seja entendido apenas como um objeto, mas como um elemento com valor agregado que representa as informações contidas no objeto.

Outra questão que pode ser discutida é o engajamento de instituições públicas nos projetos dos jornais, de modo a oferecer parcerias aos veículos de comunicação, possibilitando assim que estes recursos não fiquem disponíveis somente para assinantes, mas envolvam um número maior de usuários.

Referências

ASSOCIAÇÃO NACIONAL DE JORNAIS. **Cronologia dos jornais no Brasil**. Brasília, 2008. Disponível em: <<http://www.anj.org.br/cronologia>>. Acesso em: 22 maio 2014.

ASSOCIAÇÃO NACIONAL DE JORNAIS. **Jornais com maior circulação no Brasil**. Brasília, 2008. Disponível em: <<http://www.anj.org.br/maiores-jornais-do-brasil-2/>>. Acesso em: 15 maio 2015.

ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, p.15-27, 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>>. Acesso em: 28 mai. 2014.

BRASIL. Ministério da Cultura. Lei Federal de Incentivo à Cultura. Restabelece princípios da Lei nº 7.505, de 2 de julho de 1986, institui o Programa Nacional de Apoio à Cultura (Pronac) e dá outras providências. 1991. Disponível em <http://www.planalto.gov.br/ccivil_03/leis/18313cons.htm>. Acesso em: 23 nov. 2014.

BIBLIOTECA NACIONAL DIGITAL BRASIL. **Apresentação**. 2015. Disponível em: <<http://bndigital.bn.br/sobre-a-bndigital/?sub=apresentacao/>>. Acesso em: 26 abr. 2015.

BORGES, Maria Manuel Marques. Biblioteca digital: materialização e utopia. **Revista da Faculdade de Letras Ciências e Técnicas do Patrimônio**, Porto, v. 2, n. 1, p.653-664, 2003. Disponível em: <<http://ler.letras.up.pt/uploads/ficheiros/2936.pdf>>. Acesso em: 28 maio 2014.

CHOO, Chun Wei. **A organização do conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. 2. ed. São Paulo: SENAC, 2006.

DODEBEI, Vera. Patrimônio e Memória Digital. **Morpheus**, Rio de Janeiro, v. 5, n. 8, p.1-15, 2006. Disponível em: <<http://seer.unirio.br/index.php/morpheus/article/view/4759/4250>>. Acesso em: 4 maio 2015.

ESTADÃO. **Acervo**. São Paulo, c2007. Disponível em: <<http://acervo.estadao.com.br/faq/>>. Acesso em: 11 nov. 2014.

ESTADÃO. **Conversor de Valores**. São Paulo, c2007. Disponível em: <<http://acervo.estadao.com.br/noticias/acervo,como-funciona-o-conversor-de-valores,581,0.htm>>. Acesso em: 11 nov. 2014.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**. Brasília, v.35, n.1, p. 25-30, jan./abr. 2006. Disponível em: <<http://www.scielo.br/pdf/ci/v35n1/v35n1a03.pdf>>. Acesso em: 7 jun. 2014.

FOLHA DE SÃO PAULO. **Acervo Folha**. São Paulo, [200-]. Disponível em: <<http://acervo.folha.com.br/>>. Acesso em: 6 jun. 2014.

FOLHA DE SÃO PAULO. **Acervo Folha. Sobre o acervo**. São Paulo, [200-]. Disponível em: <<http://acervo.folha.com.br/acervo/>>. Acesso em: 06 jun. 2014.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Manifesto da IFLA para bibliotecas digitais**. 2012. Traduzido por Hanna Gledyz e Emília Sandrinelli. Disponível em: <<http://biblioo.info/wp-content/uploads/2012/11/Manifesto-IFLA.pdf>>. Acesso em: 4 maio 2014.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados:

revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, p.41-52, jan/abr. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n1/a05v31n1.pdf>>. Acesso em: 9 jun. 2014.

O GLOBO. **Acervo O Globo**. Rio de Janeiro, 2011. Disponível em: <<http://acervo.oglobo.globo.com/>>. Acesso em: 12 jun. 2014.

PALACIOS, Marcos; MIELNICZUK, Luciana; BARBOSA, Suzana et al. **Um mapeamento de características e tendências no jornalismo online brasileiro e português**. Trabalho apresentado no XXV Intercom. Salvador, 2002. Disponível em: <http://www.facom.ufba.br/jol/pdf/2002_palacios_mapeamentojol.pdf>. Acesso em: 06 jun. 2014.

PALACIOS, Marcos et al. Um mapeamento de características e tendências no jornalismo online brasileiro. In: INTERCOM, 25., 2002, Salvador. **Anais...** Salvador: Intercom, 2002. p. 1-16. Disponível em: <http://www.intercom.org.br/papers/nacionais/2002/Congresso2002_Anais/2002_NP2PALACIOS.pdf>. Acesso em: 22 mai. 2015.

PEREIRA, Fernanda. **Avaliação de usabilidade em bibliotecas digitais: um estudo de caso**. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte, 2011. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8LAKHD/dissertacao_pdf.pdf?sequence=1>. Acesso em: 28 maio 2014.

SANDUSKY, Robert J. *Digital Library Attributes: Framing Usability Research*. In: *Proceedings of the Workshop on Usability of Digital Libraries at Joint Conference On Digital Libraries*, p. 35, 2002. Disponível em: <<http://www.ucl.ac.uk/annb/docs/Sandusky35.pdf>>. Acesso em: 6 jun. 2014.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 08 jun. 2014.

SAYÃO, Luis Fernando. Afinal, o que é biblioteca digital? **Revista Usp**, São Paulo, v. 1, n. 80, p.6-17, 2009. dez./fev. Disponível em: <<http://eprints.rclis.org/14675/1/biblioteca-digital.pdf>>. Acesso em: 28 maio 2014.

TAKAHASHI, Tadao (Org.). **Sociedade da Informação no Brasil**: livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/434/1/Livro%20Verde.pdf>>. Acesso em: 25 maio 2014.

Evaluation of information retrieval in digital collections of newspapers

This work evaluates the Information Retrieval System of three Brazilian newspapers that provide faithful reproductions of daily publications that made history both in the Brazilian and international societies. For the evaluation of digital collections, we used a methodology developed by adapting Sandusky's (2002) method for digital libraries.. The analysis of the results showed that digital newspaper collections require improvement in the processes of information retrieval. The definition of metadata needs to contemplate, besides the description of the contents, associative relationships that take into account the changes in spelling implemented during the period in which the contents of the newspapers were developed, due to the historical context of the documents. Interoperability standards, preservation and open access initiatives are also indispensable. The conclusion shows the challenges of digital collections as important sources of information that allow, in addition to access to primary documents, the preservation of the cultural and historical heritage of a nation.

Key words: Digital collection. Newspaper. Information Retrieval. Scanning documents. Digital libraries.

Recebido em: 14/12/2014

Aceito em: 15/06/2015



1 DUGUID, Paul. **Report of the Santa Fe planning workshop on distributed knowledge work environments: digital libraries**. Ann Arbor: University of Michigan, 1997.