

Grupo de Trabalho 8: INFORMAÇÃO E TECNOLOGIA

UMA ARQUITETURA GENÉRICA PARA SISTEMAS DE BIBLIOTECA DIGITAL COMO PRETEXTO PARA CRIAÇÃO DE UMA AGENDA DE PESQUISA

A GENERIC ARCHITECTURE FOR DIGITAL LIBRARY SYSTEMS AS A PRETEXT FOR CREATING A RESEARCH AGENDA

Luís Fernando Sayão

Comissão Nacional de Energia Nuclear / Centro de Informações Nucleares
e-mail: lsayao@cnen.gov.br

Resumo

As bibliotecas digitais são sistemas bastante complexos que integram um conjunto de serviços voltados para a captura, a catalogação, o armazenamento, a busca, a recuperação, a preservação e a gestão da informação digital. Cada uma das diferentes concepções de biblioteca digital adota um modelo e uma arquitetura diferente. O presente trabalho discute os conceitos e os princípios básicos envolvidos nos projetos e na arquitetura de sistemas de bibliotecas digitais, ao mesmo tempo em que propõe uma arquitetura genérica para esses sistemas. A partir dessa arquitetura extrai-se uma lista de temas de pesquisa para a área. Esses temas têm como objetivo cobrir as principais questões da área de bibliotecas digitais.

Palavras-chave

Arquitetura de bibliotecas digitais; Agenda de pesquisa; Serviços e funções.

Abstract

Digital libraries are very complex systems that integrate a range of services for capturing, cataloguing, storing, searching, retrieval, preservation and management of digital information. Each one of the different conceptions of digital libraries adopts a different architecture and model. This paper discusses the basics concepts and principles involved in design and architecture of digital libraries, while proposes a generic architecture for them. From this generic architecture a list of research topics is extracted. These research topics are intended to cover the main issues of digital library area.

Keywords

Digital library architecture; Research agenda; Services and functions.

Introdução

As bibliotecas digitais representam nova infra-estrutura e ambientação para serviços de informação que vêm se desenvolvendo a partir da integração de saberes de várias disciplinas, especialmente da Ciência da Informação (CI) e da Biblioteconomia e das tecnologias de informação e de comunicação (TIC). As diferentes filosofias subjacentes a essas disciplinas têm exercido influência marcante no desenvolvimento das bibliotecas digitais, moldando um conceito expandido de biblioteca que está destinada a se tornar parte essencial da infra-

estrutura de informação do século XXI. Por conta disso, as bibliotecas digitais tornam-se foco instigante de estudos e pesquisas que têm reflexo em muitas outras áreas.

Essencialmente, uma biblioteca digital é um sistema *on-line* que proporciona acesso a uma grande variedade de conteúdos, e disponibiliza elenco diversificado de serviços voltados para gestão da biblioteca e de seus estoques informacionais e para a disseminação e o acesso a esses estoques. Os conteúdos gerenciados pela biblioteca podem incluir virtualmente qualquer tipo de material em formato digital, assim como bases de dados de periódicos, de artigos e de resumos e ainda descrições de objetos e de coleções físicas. Apesar do desafio de integrar mídias e conteúdos diferentes, cabem aos serviços de uma biblioteca digital, fundamentalmente, os mesmos papéis exigidos para a formação das coleções tradicionais presentes em bibliotecas físicas, ou seja, seleção, aquisição, descrição, armazenamento e preservação. Porém, a natureza digital das informações, seu grau de heterogeneidade e de granularidade, sua fragilidade intrínseca e o fato de não estarem concentradas num único lugar, impõem alguns desafios importantes aos pesquisadores e aos profissionais da informação.

As bibliotecas digitais são sistemas complexos e sofisticados, assim como são as bibliotecas tradicionais. Somados a essa complexidade, outros olhares menos técnicos se sobrepõem – como os olhares da ética, da privacidade e da economia – e ganham relevância na medida em que as bibliotecas digitais deixam de ser meros sistemas de informação, no sentido tradicional, e se tornam ambientes de conhecimento e de aprendizado. Esses novos sistemas não apóiam somente a busca por informação, mas são também ambientes colaborativos onde os usuários se comunicam, compartilham informações e desenvolvem trabalhos. (LIEW, 2009).

O desenvolvimento de sistemas de bibliotecas digitais e de todas suas vertentes, como repositórios institucionais e temáticas, museus e arquivos, por sua complexidade e amplitude, requer compreensão exata de todas as variáveis envolvidas – usuários, serviços, conceitos, tecnologias, padrões, conteúdos e representações – e do relacionamento entre eles. Nessa direção, o objetivo desse estudo é apresentar uma arquitetura genérica capaz de acomodar, em termos de modelo conceitual, os níveis heterogêneos de usuários, conteúdos e serviços que interagem mutuamente no âmbito de uma biblioteca digital aberta e interoperável.

A partir desse entendimento comum, passível de compartilhamento, sobre o domínio da aplicação, propõe-se a criar um exercício onde as lacunas e as necessidades de pesquisa e de aprofundamentos poderão ser identificadas, contextualizadas e expressas na forma de temas para uma agenda de pesquisa. Esses temas são “normalizados” tendo como fundamentos os assuntos de artigos básicos da área, principalmente, os publicados no final da década passada e início da atual pelos autores que impulsionaram as pesquisas em bibliotecas digitais; as agendas de pesquisas publicadas e, especialmente, os sumários do periódico D-lib (<<http://www.dlib.org>>), reconhecido, consensualmente, como o título mais importante da área de bibliotecas digitais.

Princípios

Não obstante as inúmeras vertentes em implementação de bibliotecas digitais hoje conhecidas, há nítido consenso em torno de algumas diretrizes comuns que orientam os projetos de arquiteturas de bibliotecas digitais. Seguem alguns desses princípios, reescritos a partir do enunciado de Pandey (2003) e Payette (1998):

- **Orientado para serviços** – uma arquitetura para bibliotecas digitais deve ser voltada para os serviços que a biblioteca deverá oferecer e para as ferramentas necessárias à entrega adequada desses serviços aos usuários.
- **Arquitetura aberta** – uma arquitetura para bibliotecas digitais deve ter suas funcionalidades particionadas em conjuntos de serviços bem definidos e estar baseada em padrões e protocolos abertos, que ofereçam suporte a interoperabilidade entre sistemas heterogêneos e distribuídos.
- **Ferramentas para gestão** – a arquitetura deve prever ferramentas para a gestão técnica e administração da biblioteca digital, incluindo gestão das coleções e controle de direitos.
- **Escalabilidade** – a arquitetura deve ser robusta, escalável e confiável de forma a suportar uma taxa alta e continuamente crescente de transações provenientes de um universo heterogêneo, em termos de *background* e de necessidade de informação de usuários.
- **Preservação** – a arquitetura deve assegurar acesso persistente às coleções da biblioteca digital, ao mesmo tempo em que garante a integridade de seus conteúdos, tratando de questões, tais como: nomes persistentes, arquivamento digital confiável e preservação digital.
- **Privacidade** – a arquitetura deve ser sensível às questões de privacidade – e de outras questões éticas – e dar suporte ao acesso anônimo e customizado aos recursos da biblioteca.
- **Aderência a padrões** – a arquitetura deve representar uma aderência flexível e prática aos padrões e protocolos pertinentes, assegurando a interoperabilidade e a extensibilidade.
- **Modularidade** – a arquitetura deve representar uma combinação de novas tecnologias e de tecnologias legadas que podem interagir em diferentes níveis.
- **Suporte ao cliente** – a arquitetura deve prover suporte a um universo variado de usuários: usuário-final com diversos graus de habilidade e conhecimento sobre o sistema e sobre a organização da informação; profissionais da informação e administradores das coleções e da biblioteca como um todo.
- **Federação** – permitir a agregação de coleções fisicamente distribuídas.
- **Distribuição** – permitir a gestão de conteúdos (coleções digitais) e serviços que se encontram fisicamente distribuídos.

Arquitetura e mecanismo

Um sistema de biblioteca digital é a concretização de uma arquitetura que se viabiliza através de uma infra-estrutura de *software*, de *hardware* e de rede de computadores. Dado um contexto político, econômico e técnico, existe amplo espaço de possíveis sistemas, arquiteturas e mecanismos componentes para construção de bibliotecas digitais. Porém, a natureza dinâmica e aberta que o futuro espera dessas bibliotecas argumenta contra os projetos de arquiteturas mais rígidas. As arquiteturas atuais devem ser receptivas a novos componentes e acomodar naturalmente tipos inéditos de bens de informação e serviços, assim como atender às novas demandas que irão surgir no decorrer do tempo. (SCHAUBLE; SMEATON, 1998)

Em termos de arquitetura técnica, as bibliotecas digitais são diferentes dos sistemas de OPAC (sigla em inglês para catálogos *on-line* de acesso público) que têm infra-estrutura formada por sistemas monolíticos e autocontidos. As bibliotecas digitais são compostas por sistemas e por recursos distribuídos conectados via rede e integrados, sob o ponto de vista do usuário, geralmente, por interfaces Web. Isso exige olhar unificador, infra-estrutura compartilhada que objetive o desenvolvimento da biblioteca digital como um todo coerente, ao invés de plêiade de serviços isolados. (SULEMAN; FOX, 2001). Espera-se, também, que as novas infra-estruturas possam apoiar a inserção das bibliotecas digitais no mundo do comércio eletrônico e que ofereçam bases tecnológicas para que se caminhe para uma nova economia da informação baseada em interoperabilidade entre os sistemas e em novos padrões de controle dos direitos que sejam mais harmônicos com os paradigmas atuais de acesso e de uso da informação eletrônica.

A arquitetura proposta é uma solução abstrata – mais próxima de um modelo conceitual – voltada para a descrição e a compreensão dos relacionamentos entre as diversas entidades presentes num ambiente de uma biblioteca digital aberta e interoperável. O modelo consiste de um núcleo mínimo de conceitos e de relacionamentos, independentes de tecnologias, de padrões específicos e de estratégias de implementação. (AGER, 1999; CANDELA *et al.*, 2006). O modelo está definido, conforme a Figura 1 mostra, em três camadas.

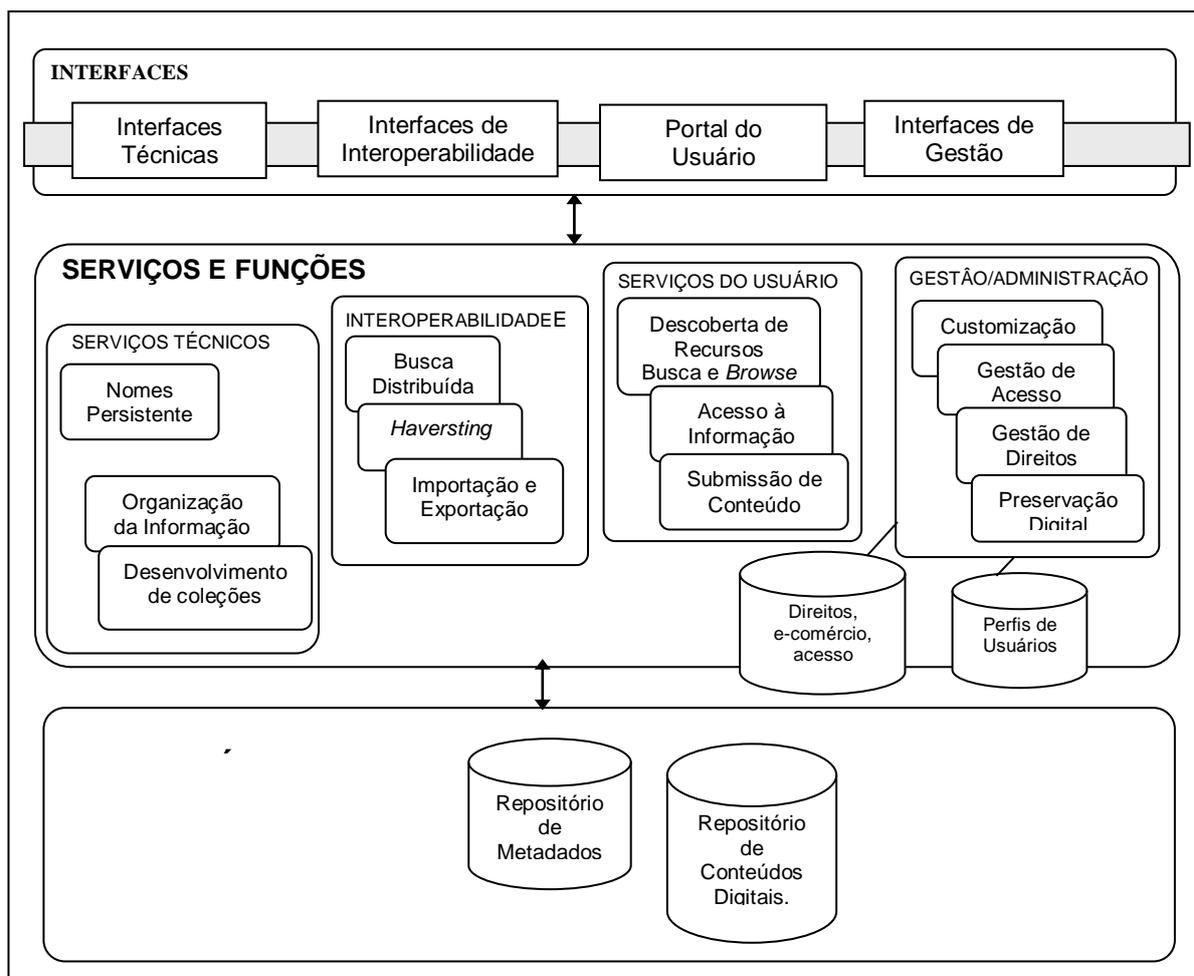


Figura 1 – Arquitetura genérica para um sistema de biblioteca digital

1. **Repositório de conteúdos** – onde estão armazenados os itens de conteúdo agrupados por coleções e os metadados relativos aos itens e às coleções.
2. **Camada de serviços** – camada ou *middleware* responsável pela gestão da biblioteca e por suas funcionalidades, incluindo a gestão de conteúdos.
3. **Interfaces** – representa a camada mais elevada da arquitetura. Dispositivos através dos quais as pessoas e sistemas interagem com o sistema de biblioteca digital. Apresenta especificidades para diferentes grupos de usuários: usuário-final, gerentes, profissionais da informação e outros sistemas.

Temas de pesquisa

As arquiteturas atuais não parecem adequadas para oferecer as funcionalidades exigidas pelos avanços que se espera para as bibliotecas digitais. É necessário que a pesquisa explore novas propostas de arquitetura e de modelos:

- Arquitetura para bibliotecas digitais baseada em componentes.
- Arquitetura multicamadas para bibliotecas digitais.

- Arquiteturas abertas voltadas para novos ambientes informacionais.
- Arquiteturas escaláveis e descentralizadas.
- Arquitetura de bibliotecas híbridas.
- Convergência de ambientes virtuais de aprendizado e bibliotecas digitais.
- Integração de ambientes de conhecimento: biblioteca, museus e arquivos digitais.
- Inserção das bibliotecas digitais no comércio eletrônico.
- Linguagem genérica para descrever bens de informação e serviços.
- Modelos de referência para bibliotecas digitais.
- Requisitos para avaliação de bibliotecas digitais.

Conteúdos digitais

A infra-estrutura computacional é meramente uma casca vazia para a criação de bibliotecas digitais. A substância que as bibliotecas digitais concretamente disponibilizam são os conteúdos informacionais intermediados pelos serviços de informação da biblioteca.

Grande parte dos primeiros desenvolvimentos de bibliotecas digitais estava concentrada em materiais que mantêm analogia direta com alguma forma física. Esses materiais, grosso modo, são representados como simples arquivo estruturado de computador. Entretanto, as bibliotecas digitais podem ir além desses objetos digitais mais simples. De fato, elas podem incluir tudo que pode ser representado em formatos digitais. Esse fato expressa o desejo do usuário que ao se referir a um item de uma biblioteca digital o faz pensando num nível mais alto de abstração do que o de um simples arquivo de computador. O meio digital permite novos tipos de objetos – ou itens – de biblioteca, tais como: *softwares*, simulações, animações, filmes, eslaides *show* e trilhas sonoras. A própria computação engendrou gêneros próprios de objetos, como as bases de dados, planilhas eletrônicas e hipertextos, entre outros. Computadores e redes sustentam o desejo – real ou induzido – por um contínuo aparecimento de tipos inéditos de informação digital. Para Arms (2000), mesmo os mais simples objetos podem vir em muitas versões e replicados muitas vezes.

Objeto digital de conteúdo

A entidade básica encontrada num sistema de biblioteca digital é o objeto digital de conteúdo – ou objeto digital – que constitui abstração através da qual a informação é armazenada, acessada, disseminada e gerenciada no domínio do sistema. Essa abstração utilizada pela arquitetura parte do pressuposto de que o conteúdo no ambiente de uma biblioteca digital é mais do que uma simples cadeia de bits (*bit stream*). Ele é um envoltório que agrega uma ou mais cadeias de bits que determinam a semântica e o comportamento do objeto.

Este conceito elaborado do objeto digital se justifica pela necessidade de refletir a sofisticação de um ambiente em contínua mudança, com demandas sofisticadas e alto grau de especialização, o qual exige que a idéia de objeto digital transite da condição de “item único” em direção a uma estrutura rica que comporte objetos informacionais que sejam multimídias, multilíngües, anotados, etc. Nessa direção, os serviços básicos de uma biblioteca digital devem acolher novo modelo de objeto digital capaz de representar a ampla variedade de documentos, que são, caracteristicamente, multiversões e heterogêneos em termos de conteúdo, mídia e estrutura.

A expansão do conceito de objeto digital possibilita o modelamento do que se conhece como “objeto complexo”, significando que o objeto é formado por um pacote de recursos intrinsecamente agregados que incluem as diferentes percepções, manifestações e representações do mesmo conteúdo intelectual. Por exemplo, um item depositado num repositório é composto por uma ou mais manifestações desse item (um arquivo *Portable document format* ou PDF, um documento Word, etc.). Cada uma das manifestações pode ter registros de metadados separados associados a ela, sendo que os metadados representam tipo particular de visão do objeto. A agregação das diversas percepções e representações de conteúdos e dos metadados que formam um objeto digital complexo é empacotada segundo padrões do tipo *Metadata Encoding and Transmission Standard* (METS) e *Moving Picture Experts Group* (MPEG-21). Esses padrões simplificam a forma pela qual esse tipo de objeto pode ser depositado, gerenciado e recuperado no âmbito da biblioteca digital e de outros repositórios digitais, conforme Powell (2005). Dessa forma, se estabelece que a arquitetura deva proporcionar meios técnicos de manipular os pacotes.

A arquitetura ora discutida é composta por dois tipos de objetos digitais: itens e coleções.

Objeto digital de conteúdo: itens de conteúdo ou itens

São os objetos fundamentais no contexto de uma biblioteca digital, correspondendo diretamente ao acervo físico – como livros e periódicos – custodiados por uma biblioteca tradicional. Ainda como nas bibliotecas convencionais, onde cada recurso é identificado por esquemas padronizados locais ou globais, como o *International Standard Book Number* (ISBN), o *International Standard Serial Number* (ISSN) ou números locais de registro, os recursos de uma biblioteca digital são distinguidos inequivocamente de outros por esquemas de nomes ou identificadores persistentes.

Os metadados cumprem papel de vital importância em todos os fluxos técnicos da biblioteca relacionados aos seus estoques informacionais: os itens são descritos em termos de conteúdo por esquemas de metadados descritivos e em termos de dependência de *hardware* e *software* por metadados técnicos; a gestão dos itens é apoiada pelos metadados administrativos; e os metadados estruturais representam os relacionamentos entre um item e as partes que os compõem.

Objeto digital de conteúdo: coleções

No contexto da arquitetura proposta, os conteúdos são colocados disponíveis coletivamente na forma de coleções. Uma coleção é interpretada como agregação de um ou mais itens, desenvolvida segundo algum critério, política ou missão de significância para uma dada comunidade-alvo. Podem existir diversos tipos de coleções, incluindo coleções de recursos textuais, de imagens, de dados, de objetos de aprendizagem e assim por diante. Entretanto, as coleções podem ser heterogêneas, significando que podem conter diferentes tipos de itens; e podem existir também coleções de metadados sobre outras coleções. Em termos de localização física, as coleções podem estar armazenadas localmente, em máquinas servidoras da instituição, em servidores externos ou podem, ainda, estar distribuídas por várias máquinas internas e externas. Um item pode pertencer a mais de uma coleção. Para tal, as informações sobre um item individual, bem como sobre a coleção como um todo, são mantidas no nível da coleção. (POWEL; LYON, 2001).

É interessante notar que em algumas propostas como a da arquitetura ADEPT (JANÉE; FREW, 2002) é definido o objeto de conteúdo biblioteca, significando um conjunto de coleções.

Temas de pesquisa

Um tópico de pesquisa dos mais importantes está relacionado com a compreensão dos objetos que estão numa biblioteca digital. As bibliotecas digitais armazenam e disseminam qualquer informação que pode ser representada em forma digital. Um desdobramento direto disso é que os problemas de representar e manipular informação são variados e sutis.

- Armazenamento e intercâmbio de objetos digitais complexos.
- Conceito de coleção distribuída.
- Descrição de objetos digitais.
- Gestão de objetos digitais em repositórios institucionais.
- Gestão de coleções digitais distribuídas.
- Modelos de objetos para bibliotecas digitais.
- Objetos digitais museológicos.
- Objetos digitais de arquivo.
- Padrões para estruturação e empacotamento de objetos digitais complexos.
- Sistema de repositório para objetos digitais complexos.
- Uso do METS para arquivamento de documentos eletrônicos.

Tipos de informação

A arquitetura define três tipos de informações externamente visíveis: metadado, consulta e conjunto de resultados.

Metadado

É informação sobre itens ou sobre coleções. São instrumentos necessários ao acesso, à estruturação e à gestão dos objetos digitais e formam uma plataforma sobre a qual são construídos os serviços essenciais da biblioteca. Os registros de metadados sobre um item particular descrevem seu conteúdo intelectual, sua estrutura, seus direitos associados, suas manifestações, suas representações (por exemplo, miniaturas associadas), etc. Os metadados sobre uma coleção descrevem a coleção como um todo, incluindo origem, distribuição geográfica, lei de formação, etc. Eles têm a função de registrar a agregação de itens que possuem algum grau de vinculação constituindo uma coleção. A agregação definida pelos registros de metadados de coleção pode ser de natureza administrativa, semântica ou pessoal. Observa-se que um item pode estar agregado a uma ou mais coleções. Os registros de metadados de coleção servem, ainda, como base administrativa para a gestão de importação e de atualização de itens. É possível, também, definir o registro de metadados de relacionamento, que tem a função de descrever o relacionamento entre os metadados dentro do repositório. O exemplo mais comum é a descrição da relação hierárquica entre um registro de coleção e os registros de item que ele agrega. (LAGOZE *et al.*, 2002).

Consulta e conjunto de resultados

Uma consulta é um predicado direcionado a uma ou mais coleções de objetos pertencentes a uma ou mais bibliotecas. Um conjunto de resultados, por sua vez, é um conjunto – que pode ser vazio – de identificadores de itens que satisfazem uma consulta. Os identificadores podem estar associados a outros metadados que descrevam os itens recuperados e permitem acesso, quando possível, aos conteúdos.

Principais componentes

Repositórios

Numa definição direta, um repositório é um sistema de armazenamento acessível via rede de computadores, no qual objetos digitais, incluindo registro de suas representações sob a forma de metadados, podem ser armazenados tendo em vista a recuperação e o acesso. (KAHN; WILENSKY, 2006). A infra-estrutura proposta estabelece dois tipos distintos de repositórios: repositório de metadados e repositório de conteúdos digitais, cujos respectivos escopos podem ser assim descritos.

Repositório de metadados

No caso do repositório de metadados, trata-se de um componente central da arquitetura que provê facilidades para o armazenamento de registros de metadados e do relacionamento entre eles, criando a base sobre a qual se viabiliza o conjunto de serviços essenciais da biblioteca, discutido mais adiante. Conforme definido no item alusivo às interfaces do usuário, os metadados descrevem dois tipos de informação: os itens e as coleções.

É importante enfatizar que o repositório de metadados expõe os metadados nele armazenados intermediados por um conjunto de interfaces de saída definidas de forma adequada para o processamento por serviços externos ao repositório. Isso significa que os usuários – pessoas ou sistemas – não interagem diretamente com o repositório, e, sim, através dessas interfaces que dão acesso aos serviços que processam os metadados. Desses serviços, o mais destacado é o serviço de descoberta de recursos, realizado por meio de busca e por navegação por índices (*browse*).

Outro serviço fundamental é a entrada de metadados no repositório, que permite tanto a incorporação de novos registros de metadados como a atualização dos existentes. A infra-estrutura aqui proposta estipula que os metadados são incorporados ao sistema intermediados pelas interfaces de entrada, as quais podem acionar cinco mecanismos distintos que são analisados no item referente ao serviço de indexação. Os conteúdos do repositório de metadados podem ter origens diversificadas, posto que são constituídos, possivelmente, por contribuições provenientes de outros repositórios efetivadas por meio de processos automatizados, como coleta automática (*harvesting*) e agregação automática (*gathering*). Por combinar metadados de muitas coleções, o repositório de metadados pode ser considerado como uma generalização do conceito de catálogo coletivo, ou seja, catálogo que combina registros de muitas bibliotecas, conforme descrito por Lagoze *et al.* (2002).

Repositório de conteúdos digitais

Uma exigência fundamental para uma arquitetura de biblioteca digital é a definição de um dispositivo confiável e seguro para armazenamento e acesso aos conteúdos digitais. Quando se trata, como é o caso, de bibliotecas digitais cujos conteúdos são diversificados, compostos e, possivelmente, distribuídos, o repositório de conteúdos digitais exerce essa função. Para tal, ele deve ser capaz de: **(1)** dar suporte a tipos de conteúdos heterogêneos; **(2)** agregar conteúdos mesclados, possivelmente distribuídos, na forma de objetos digitais complexos; **(3)** dispor de mecanismos para gestão de acesso aos conteúdos digitais. (LAGOZE; PAYETTE, 1998).

Os serviços associados a um repositório devem oferecer mecanismos para depositar objetos às suas coleções, para armazená-los e para torná-los disponíveis ao acesso. O repositório deve ser suficientemente extensível para acomodar variedades inéditas de objetos e apoiar novos tipos de disseminação e de mecanismos para gestão de direitos.

Formalmente, considera-se que um objeto digital está contido num dado repositório se seu identificador persistente (*Uniform Resource Name / URN; Digital object identifier / DOI; Handle, etc.*) é resolvido no endereço desse particular repositório, implicando que o acesso ao objeto só se viabiliza via requisição de serviço feita a esse repositório, conforme afirmação dos autores supracitados. Os repositórios proporcionam aos usuários acesso aos objetos digitais por eles armazenados segundo termos e condições estabelecidas pelo depositante e / ou por políticas específicas de acesso do repositório em questão. (KAHN; WILENSKY, 2006).

Temas de Pesquisa

- Armazenamento persistente de pacotes METS de informação.
- Armazenamento de metadados para a otimização da recuperação de informação.
- Projeto de repositórios seguros para bibliotecas digitais.
- Sistema de armazenamento para objetos digitais complexos.
- Arquitetura de armazenamento distribuído para biblioteca digitais.

Interfaces

Toda a interação do sistema com os atores envolvidos – usuários, profissionais da informação e outros sistemas – será realizada por intermédio de interfaces, que são janelas através das quais os serviços são acionados e operacionalizados. A arquitetura identifica quatro conjuntos distintos básicos de interfaces – interfaces do usuário, interfaces técnicas da biblioteca, interfaces de gestão e interfaces de interoperabilidade –, detalhadas a seguir.

Interfaces do usuário

Disponibilizam pontos de entrada, centrados no ser humano e usando padrões comuns da Web, para as funcionalidades oferecidas ao usuário-final da biblioteca. Por conseguinte, o usuário-final acessa as coleções digitais e os demais serviços da biblioteca intermediados por esses dispositivos, que, coletivamente, serão chamados de portais ou *gateways* do usuário. As interfaces do usuário apresentam visão organizada e coerente de recursos e serviços consubstanciados na forma de aplicações Web. A idéia básica das interfaces do usuário é que sejam capazes de filtrar a complexidade tecnológica do sistema e, ao mesmo tempo, permitam

ao usuário usufruir diferentes visões da biblioteca e de suas coleções de acordo com suas preferências e habilidades. Isso significa que o mesmo conjunto de serviços essenciais da biblioteca estará disponível por meio de várias interfaces, e, por conseguinte, poderá ser apresentado de maneiras distintas. Nessa direção, o *design* das interfaces apresenta alto grau de customização. Um exemplo de formulações de interfaces são os portais especializados, dirigidos para comunidades específicas de usuários, como pesquisadores e estudantes. Outro exemplo são os portais personalizados que permitem graus diferenciados de customização de recursos e serviços, baseados, por exemplo, em dados recuperados de servidores de perfil de usuário. (LAGOZE *et al.*, 2002).

Interfaces técnicas da biblioteca

É o conjunto de interfaces que facilitam a execução por parte dos profissionais da informação, das funções técnicas da biblioteca, traduzidas em serviços, tais como indexação, catalogação, classificação e atribuições de nomes persistentes. Em resumo: essas interfaces coletivamente constituem as ferramentas de intermediação entre sistema e profissionais técnicos de informação. Devem dispor de mecanismos que permitam ajustes segundo o grau de *expertise* dos profissionais em relação ao sistema e apoiem a eficiência dos processos, como no caso de se utilizar ferramentas de *workflow*.

Interfaces de gestão da biblioteca

São as interfaces através das quais os gerentes da biblioteca podem executar as políticas da instituição bem como ajustar as opções de funcionamento para a biblioteca como um todo ou para coleções e itens específicos. Exemplificando: política de indexação, gestão de acesso, padrões de preservação digital.

Interfaces de interoperabilidade

São as interfaces por meio das quais o sistema intercambia informações com sistemas externos segundo padrões e protocolos bem estabelecidos. Essas interfaces podem oferecer informações a sistemas externos, bem como fazer uso de informações desses sistemas para ampliar seus serviços. Um exemplo ilustrativo desse tipo de dispositivo é a interface *Open Archive* para provedores de dados, que expõe metadados da biblioteca para coleta automática por outros sistemas; ou a busca distribuída por várias bibliotecas segundo o protocolo Z39.50.

Temas de pesquisa

Desenvolver as formas de intermediação do usuário com a informação num ambiente conectado por redes é um assunto de significativa importância para uso pleno das bibliotecas digitais, principalmente, pela ampla audiência que se espera para elas num futuro previsível. Um enfoque bastante sensível para os atuais sistemas é a questão da personalização, que envolve multiplicidade de variáveis, mas, que sem dúvida, passa pelo aperfeiçoamento das interfaces.

Um tema subjacente à questão das interfaces é o que trata das questões comportamentais. Para desenvolver bibliotecas digitais usáveis e melhorar os projetos de sistema, os pesquisadores têm reiteradamente considerado o comportamento e as exigências

dos usuários em diferentes contextos, incluindo os ambientes acadêmicos, escolares, governamentais, departamentais, de negócios e de entretenimento. Para Shiri (2003), um dos maiores desafios nos estudos está associado com as propostas metodológicas e as técnicas de coleta de dados.

A acessibilidade – capacidade que uma biblioteca digital dispõe para facilitar sua localização e sua consulta por todos os usuários, independentemente de suas limitações físicas e técnicas – é investigada no âmbito das interfaces, segundo Céspedes (2006). Os seguintes tópicos são pontos de interesse para a área:

- Competência informacional de usuários de bibliotecas digitais.
- Biblioteca digital como suporte ao aprendizado a distância.
- Estudos empíricos da interação usuários e bibliotecas digitais.
- Geração automática de interfaces para interação personalizada.
- Modelos de usuários.
- Padrões de acessibilidade para bibliotecas digitais.
- Parâmetros para avaliação de interfaces de bibliotecas digitais.
- Usabilidade, acessibilidade e aceitação de bibliotecas digitais.

Serviços essenciais

A arquitetura proposta define um conjunto básico de serviços de biblioteca digital cujas interações estão representadas na Figura 1. Esses serviços, chamados de serviços essenciais, formam um núcleo que traduz as funcionalidades básicas de um sistema de biblioteca digital. Esses serviços são bastante similares aos oferecidos pela maioria das bibliotecas convencionais – exceto pelas idiossincrasias da natureza digital da informação – e compõem uma lista clássica: desenvolvimento de coleções, tratamento técnico, descoberta de recursos, acesso, gestão, preservação e interoperabilidade. Como a arquitetura se configura como arquitetura aberta, em que as funcionalidades estão expostas através de serviços baseados em protocolos, outros serviços podem ser incorporados para enriquecer esse núcleo essencial de funcionalidades.

Temas de pesquisa

- Padrão *plug-and-play* para incorporação de novos serviços.
- Autodescrição, auto-registro e autoconfiguração de serviços e coleções.

Os serviços essenciais são segmentados em três tipos: serviços técnicos da biblioteca; serviços do usuário; serviços de gestão da biblioteca e serviços de interoperabilidade.

Serviços técnicos da biblioteca

Os serviços técnicos da biblioteca compreendem os serviços biblioteconômicos necessários ao tratamento técnico da informação. São eles operados por meio das interfaces voltadas para os profissionais de informação, chamadas interfaces técnicas, como visto. São os seguintes os serviços essenciais enquadrados nesta categoria:

1. Serviço de atribuição de nomes persistentes

Em consonância com o pensamento de Powell (2005), os nomes constituem parte essencial da descrição dos objetos digitais no contexto da infra-estrutura proposta. Eles são tão importantes numa biblioteca digital quanto o ISBN é para uma biblioteca tradicional. Para que um sistema de bibliotecas funcione com um grau aceitável de integridade, esses nomes devem ser persistentes e globalmente únicos, e não devem estar ligados a uma localização específica ou a processos ou tecnologias. Nessa direção, a biblioteca digital deve dispor de serviço capaz de identificar univocamente e de forma persistente todos os recursos informacionais gerenciados por ela ou que estejam a ela associados, tendo como propósito a citação, a recuperação da informação, o *link* entre objetos e a gestão de direitos. É importante observar que essa exigência não está limitada somente aos objetos digitais (itens e registros de metadados), mas pode incluir, também, objetos físicos (autores, livros, etc.) e ainda os objetos conceituais, como os termos de um vocabulário.

Essa função é de grande importância para as bibliotecas digitais, posto que entre a descoberta de um recurso informacional e o acesso a ele, o usuário-final deve estar de posse de informações suficientes para saber como requisitar precisamente uma instância particular desse recurso. Para tal, é necessário que o identificador do recurso seja traduzido ou resolvido em sua localização. Para cada identificador, o serviço de descoberta de recursos deve responder com, pelo menos, um conjunto de metadados do objeto identificado que, no mínimo, inclua sua localização. Exemplificando: um exemplar de um livro impresso cujo identificador é seu ISBN tem como resposta – ou é resolvido – por sua localização na estante; por outro lado, o DOI de um documento eletrônico é resolvido em seu *Unified Resource Locator* (URL), que pode se concretizar num *link* para o recurso.

O serviço de atribuição de nomes pode ser um módulo do próprio sistema ou uma facilidade oferecida por alguns dos serviços externos de registro de nomes, como o DOI, *Handle System* e o *Persistent URL* (PURL). Em ambos os casos, os padrões estabelecidos devem fundamentar o serviço. (POWELL, 2005; SAYÃO, 2007).

É importante notar que a resolução do identificador persistente pode estar associada com a obtenção de informações sobre a disponibilidade de diferentes formatos do mesmo recurso – papel ou formas eletrônicas, por exemplo – e determinações sobre direitos de acesso e termos e condições associadas com instância particular do recurso. (POWELL; LYON, 2001).

Uma infra-estrutura de nomes – de âmbito nacional e internacional - é de importância fundamental para a concretização dos projetos mais abrangentes de interoperabilidade. Portanto, trata-se de área de grande interesse para o desenvolvimento de qualquer tipo de repositórios digitais. Seguem temas para estudos.

Temas de pesquisa

- Algoritmos para a geração automática de nomes persistentes.
- Adoção de esquemas de nomes persistentes em áreas específicas.
- *Links* referenciais e a nova economia da informação baseada em *pay-per-view*.
- Padrão de identificadores para as bibliotecas digitais brasileiras.

2. Serviço de indexação

O serviço de indexação, no presente contexto, é definido como componente da infra-estrutura da biblioteca digital voltada para a estruturação de mecanismos que facilitem a

descoberta de recursos. O serviço tem como princípio a assinalação de metadados descritivos aos recursos disponibilizados pela biblioteca.

A indexação tem como finalidade coletar informações sobre os objetos digitais – itens e coleções – na forma de representações familiares como são, por exemplo, os registros de catalogação na biblioteca tradicional; ou na forma de extração de dados a partir do texto completo, como são geralmente usados pelas máquinas de busca. As informações coletadas por intermédio desses processos estão organizadas em índices estruturados que permitem que dispositivos de busca respondam às solicitações de informações com precisão, revocação e eficiência; e também que o usuário navegue ou percorra (*browsing*) índices que são a representação de como a informação está organizada, por exemplo: por assunto, por autores, por departamentos, por laboratórios, etc. (LAGOZE; PAYETTE, 1998).

Na visão de Lagoze *et al.* (2002), a infra-estrutura proposta prevê a entrada de metadados descritivos pelas seguintes vias:

1. Por entrada direta ou manual – a administração da biblioteca pode autorizar um conjunto de usuários, geralmente profissionais de informação – a criar, editar e apagar registros de metadados. Os processos de indexação são assistidos pela interface técnica de indexação e por suas ferramentas, como controle de vocabulário e de autoridades, editores de metadados, seleção de esquemas e outros. Para os sistemas de bibliotecas digitais que permitem auto-submissão, como são a maioria dos repositórios temáticos e os institucionais, o usuário-final tem acesso às facilidades de entrada de dados, via interface do usuário. Entretanto, deve estar submetido aos controles necessários e à política estabelecida pelos gestores da biblioteca.
2. Por indexação automática – entendido como emprego de metodologias de extração de metadados a partir dos itens de conteúdo. Em geral, cada sistema emprega algoritmos próprios para assinalar registros de metadados que descrevem os itens de conteúdo. Porém, é cada vez mais comum que os sistemas de bibliotecas digitais empreguem programas de indexação e algoritmos produzidos por outrem. Nesse caso, o sistema pode oferecer a possibilidade de escolha do programa e, conseqüentemente, do método de indexação.
3. Por coleta automática – na qualidade de estrutura aberta, a biblioteca, através das suas interfaces de interoperabilidade, pode coletar metadados de outros repositórios aderentes a especificações apropriadas, como o protocolo *Open Archive Initiative – Protocol for Metadata Harvesting* (OAI-PMH).
4. Por incorporação em formatos padronizados – submissão de metadados codificados em formatos padronizados, como por exemplo, em XML segundo METS, ou mesmo numa planilha MS-Excel.
5. Por agregação automática – os metadados podem ser incorporados ainda por meio da combinação da extração de informações de coleções abertas ao acesso via ferramentas de busca (*web crawlers*, por exemplo); e da geração automática de metadados baseada no conteúdo extraído.

É oportuno lembrar que outros gêneros de metadados que expressam, por exemplo, as dependências tecnológicas, as restrições de uso e a estrutura do objeto, são fundamentais para a gestão dos objetos digitais, mas que, entretanto, são assinalados em outras instâncias.

Temas de pesquisa

O tratamento técnico – catalogação e indexação – são atividades caras quando desempenhadas por profissionais qualificados. Esse fato reforça a necessidade de pesquisas de métodos efetivos de indexação assistida por computador. A infra-estrutura proposta prevê o uso de ferramentas de indexação automática, modulares e independentes, selecionadas pelo grau de aderência à política de indexação da biblioteca. Considera, também, o uso de tecnologias semânticas e ontologias para redefinir em bases mais atuais a questão do significado na indexação automática. Os seguintes temas são sugeridos:

- Estudo comparativo das ferramentas de indexação automática disponíveis no mercado.
- Estruturação semântica de textos a partir de ou tendo como modelos ontologias.
- Indexação semântica baseado em ontologias.
- Indexação automática de informações não textuais.
- Métodos de indexação automática baseada em tesouros.

3. Organização da informação (para a busca por navegação por índices)

A busca direta não é a única estratégia utilizada pelos usuários para descobrir informação. O *browsing* – traduzido neste contexto por navegação por índices – é o termo geral para a exploração de um corpo de informações; é um método freqüente e efetivo para ajudar a descobrir o inesperado. Os usuários de bibliotecas digitais, de acordo com Arms (2000), comumente, lançam mão de estratégias que combinam busca e *browsing* na busca por recursos de informação.

A maioria das bibliotecas arranja as coleções por assunto com o intuito de ajudar o usuário a percorrer seus índices. Os esquemas de classificação – o Dewey, por exemplo – provê estruturação por assunto e hierárquica que pode ser usada para organizar as coleções. A forma como o sistema organiza as informações depende dos diferentes contextos em que a biblioteca opera. Por exemplo, um repositório com vertente arquivística precisa que suas informações estejam organicamente organizadas em fundos, dossiês, processos; enquanto para um repositório institucional de uma universidade, talvez seja interessante uma organização por comunidades, como laboratórios, departamentos, projetos, etc.

A infra-estrutura proposta prevê ferramentas para a organização dos conteúdos, para a definição de índices e de suas estruturas e de sistemas de navegação que possam refletir a natureza da biblioteca, de seus estoques informacionais e de suas comunidades.

Temas de pesquisa

- Arquitetura de informação para bibliotecas digitais.
- Desenvolvimento de padrões de organização de informação para repositórios tendo em vista a interoperabilidade.

4. Serviço de desenvolvimento de coleções digitais

A arquitetura oferece ferramentas técnicas e gerencias que dão suporte aos três métodos essenciais de construção de coleções digitais, resumidamente: digitalização de coleções impressas, aquisição de obras digitais originais e acesso a materiais digitais externos.

Um conjunto de parâmetros determina a política de desenvolvimento de coleções da biblioteca. Essa política define o que deve ser adquirido, o que deve ser acessado externamente e o que deve ser convertido para formatos digitais. Para tal, os gestores da biblioteca consideram variáveis, tais como: prioridades da comunidade de usuários, disponibilidade, direitos autorais, tipologia, singularidade e fragilidade dos materiais, além, é claro, das dependências técnicas e orçamentárias. Por exemplo, a digitalização pode variar da conversão retrospectiva a somente um conjunto selecionado de materiais; os formatos precisam ser os mais apropriados para as funções de acesso e preservação. Portanto, são muitas as escolhas, o que impõe decisões gerenciais críticas ao longo do processo de desenvolvimento de acervos digitais. (CARIGNAN *et al.*, 2007).

O Grupo de Trabalho em Infra-estrutura da *National Information Standards Organization* (NISO, 2007) lista os predicados que uma coleção digital deve possuir. Resumidamente, são os seguintes: possuir política explícita de desenvolvimento; ser descrita por metadados; ser sustentável ao longo do tempo; ser amplamente disponível; respeitar os direitos de propriedade intelectual; possuir mecanismos de mensuração de uso; estar inserida em contextos nacionais e internacionais.

Embora pareça, numa análise mais desatenta, atividade simples, o desenvolvimento de coleções digitais é um dos processos mais sensíveis para o sucesso da biblioteca digital. A complexidade reside especialmente em três fatores: **(1)** na necessidade das bibliotecas digitais romperem a barreira da informação puramente textual, em termos de objetos e de suas representações; **(2)** na multiplicidade de formatos e de mídias que devem ser integrados – inclusive com materiais convencionais – proporcionando visão homogênea de todo o acervo (importante notar que esse fenômeno é fortemente impactado pela taxa crescente de novas modalidades de objetos digitais que a concorrência exacerbada e a inovação tecnológica fazem brotar cotidianamente); **(3)** em como a informação deve ser organizada para o armazenamento e a para a recuperação tendo em vista um público-alvo que pode ter uma base de conhecimento específica, mas que, na maioria dos casos, será altamente diversificado em todos os sentidos. Pode-se incluir ainda nesse último item questões relativas à construção de coleções multilíngües e / ou multiculturais.

Diferentemente das coleções convencionais que têm uma sobrevivência inercial, para a sobrevivência das coleções digitais é essencial o estabelecimento de ações continuadas que assegurem sua sustentabilidade ao longo do tempo. Isso inclui intervenções para a preservação de longo prazo (analisado no item referente ao serviço de preservação das coleções digitais, mais adiante), e o desenvolvimento de modelos de custo que incorporem itens como armazenamento e crescimento da coleção.

Por fim, é necessário considerar que as coleções podem não estar concentradas num único ponto. Podem estar distribuídas em termos geográficos e de máquinas servidoras. Esse fato acrescenta um grau a mais de complexidade na gestão do desenvolvimento de coleções digitais.

Temas de pesquisa

Eis alguns temas importantes que precisam ser aprofundados:

- Aquisição automática de conteúdos digitais.
- Convergência e integração de mídias no contexto das bibliotecas digitais.
- Desenvolvimento de coleções digitais museológicas.
- Desenvolvimento de coleções digitais distribuídas.
- Desenvolvimento de coleções multilíngües / multiculturais.
- Estratégia de desenvolvimento de coleções digitais: políticas e gestão.
- Modelos de custo para desenvolvimento de coleções digitais.
- Modelos de avaliação de coleções digitais.
- Novos gêneros de objetos digitais.
- Sustentabilidade de coleções digitais.
- Padrões para a digitalização de materiais para bibliotecas heterogêneas.
- Planejamento de coleções digitais

Serviços do usuário

Os serviços do usuário são os serviços voltados para o usuário-final da biblioteca digital. São visíveis para ele através da intermediação do portal de usuários, que oferece diversas interfaces adequadas ao seu nível de especialização. São os seguintes serviços essenciais enquadrados nesta categoria: descoberta de recursos, acesso à informação e submissão de conteúdos.

Serviços do usuário: descoberta de recursos

O usuário sempre tem a expectativa positiva de que a biblioteca pode assisti-lo na busca pela informação. De longe, o serviço mais importante de uma biblioteca é esse: ajudar s usuários a localizar e recuperar a informação que deseja. No contexto das bibliotecas digitais, isso continua sendo verdadeiro, porém ainda está circunscrito a metodologias que podem ser automatizadas, indicando que algumas sutilezas do serviço somente as bibliotecas físicas podem prestar plenamente, como a interação pessoal em torno de uma busca. Entretanto, novos padrões de estudos e de modelos de usuário podem minimizar esse problema.

O objetivo do serviço de descoberta de recursos é oferecer facilidades básicas para localizar recursos informacionais, incluindo coleções, no âmbito de uma biblioteca ou de um conjunto de bibliotecas digitais. A necessidade do usuário é traduzida por uma consulta, cuja sintaxe é independente da arquitetura e deve ser especificada para incorporar as facilidades mais comuns presentes nas máquinas de buscas comerciais e da Web. Entretanto, conforme visto no item organização da informação (para a busca por navegação por índices), a busca direta é somente uma das estratégias que as pessoas usam para descobrir informações. A outra é a navegação, termo geral para denotar a exploração não estruturada no corpo das informações da biblioteca.

No contexto da arquitetura proposta, esse componente permite que qualquer item representado no repositório de metadados, como um registro de metadados, seja encontrado, pressupondo que o item inclua informações apropriadas e precisas. É importante salientar que o repositório de metadados também inclui registros de metadados em nível de coleção,

permitindo que as informações sobre coleções sejam também indexadas e estejam, portanto, disponíveis para busca. (LAGOZE *et al.*, 2002.).

O fluxo da busca inclui o recebimento de informações na forma de consulta e o retorno de uma lista de informações ordenadas sobre os objetos recuperados, na forma de registros de metadados, que inclui, minimamente, o identificador ou nome dos objetos. O serviço oferece facilidades para busca nos conjuntos de registros de metadados, como nos registros *Dublin Core* (DC). Adicionalmente, oferece mecanismos de buscas baseados nos conteúdos dos recursos ou, ainda, buscas baseadas na combinação de ambos os métodos.

A descoberta de recursos por navegação por índices depende totalmente de como a informação é organizada e das opções de exibição dessa organização para o usuário-final. Desta forma, esse componente de arquitetura disponibiliza as potencialidades fundamentais para a localização de recursos no âmbito da biblioteca e de outras unidades que têm relações de interoperabilidade implementadas com o sistema, por exemplo, via protocolo Z39.50 ou via protocolo OAI-PMH. O serviço é diretamente acessível a partir do portal do usuário, através das interfaces de descoberta de recursos, que oferece ferramentas de auxílio, incluindo níveis diferenciados de busca e reconhecimento do grau de *expertise* do usuário.

Temas de pesquisa

Localizar e recuperar informação são aspectos vitais para as bibliotecas. Buscar um item específico de informação em meio a uma volumosa coleção de textos constitui há muito tempo um desafio que os cientistas da computação e da informação, especialmente, os que trabalham na área de pesquisa conhecida como recuperação da informação (*information retrieval*) têm sistematicamente enfrentado. Mas não podemos esquecer que esses mesmos pesquisadores estão lentamente e continuamente fazendo progressos na descoberta de recursos em outros formatos que não somente textos, tais como músicas, imagens, vídeos, etc. Esse domínio constitui área ativa de pesquisas, assim como as áreas relacionadas ao processamento de linguagem natural aplicada à busca de informação. Soma-se à complexidade da questão o fato de as bibliotecas e os demais repositórios digitais estarem fisicamente distribuídos e, apesar disso, terem de apresentar visão integrada e coerente ao usuário.

Por outro lado, é necessário observar, também, que a informação digital pode estar decomposta em níveis diferenciados de granularidade, formando unidades identificáveis que podem ser recompostas formando outros objetos, como, por exemplo, um gráfico ou uma fotografia dentro de um livro. Há tantas formulações novas de objetos e de combinações inovadoras que é indispensável a criação de novos algoritmos de busca e de estruturas que possam explorar as particularidades desses objetos. (DIGITAL LIBRARIES, 2001), sem que se perca a perspectiva do usuário.

- Ambiente de interação e negociação por computador para descoberta de recursos em bibliotecas digitais.
- Aplicações de *Resource Description Framework* (RDF) na descoberta de recursos.
- Desenvolvimento de algoritmos inovadores para recuperação de objetos digitais.
- Descoberta de recursos em bibliotecas digitais distribuídas.
- Modelos de avaliação de recuperação de informação em bibliotecas digitais.
- Otimização da recuperação baseada em modelagem de usuário.
- Otimização automática de consultas em bibliotecas digitais.
- Processamento de consultas em linguagem natural.

- Recuperação de informação não textual.
- Recuperação de informação em repositórios digitais distribuídos.
- *Software* agente para recuperação de informação.

Serviços do usuário: acesso à informação

O serviço de acesso à informação é entendido, no domínio da infra-estrutura proposta, como o serviço de entrega ao usuário do recurso informacional que foi objeto de busca e de recuperação por meio de acionamento do serviço de descoberta de recursos descrito no item anterior.

O serviço de descoberta de recursos apresenta como resultado um conjunto (que pode ser vazio) de registros de metadados sobre recursos que são de provável interesse do usuário-final. Cada registro de metadados inclui, obrigatoriamente, algum tipo de identificador do item que ele descreve. É de grande importância para funcionamento consistente do sistema que o identificador obtido durante o processo de descoberta seja, ao mesmo tempo, persistente e independente de localização.

De forma independente do tipo de identificador – por exemplo, um URN ou um ISBN -, o acesso à informação desejada se efetiva quando o identificador obtido no processo de descoberta de recursos é resolvido num endereço de uma instância particular do recurso, ou ainda em informações sobre como requisitar o recurso. Por exemplo: um *link* para um artigo eletrônico, a localização na estante ou a apresentação de dados sobre a aquisição do recurso no site do editor.

No âmbito da infra-estrutura apresentada, os mecanismos de resolução podem ter suas funções complementadas com a adoção de alguns padrões já consagrados que potencializam e personalizam o serviço de acesso. Este é o caso da aderência ao padrão OpenURL, que possibilita a entrega ao usuário-final da cópia mais apropriada para ele, em termos de direitos de acesso, de custos, do acervo da instituição, etc. Para tal, o OpenURL oferece meios de codificar os dados bibliográficos do recurso como um URL e considera a identidade do usuário-final, onde ele está, à qual instituição está afiliado e que direitos de acesso possui. O sistema deve considerar, por exemplo, que para um pesquisador brasileiro vinculado a uma universidade federal que acessa um artigo eletrônico a partir de sua instituição, a cópia mais apropriada é via Portal da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e não no *site* do editor.

Temas de pesquisa

- Cópia apropriada: personalização no acesso à informação.
- Desenvolvimento de serviços de acesso à informação baseada na norma OpenURL.

Serviços do usuário: serviço de submissão de conteúdo

A internet abriu a possibilidade inédita de acesso universal a um volume continuamente crescente de dados e informações. Ao mesmo tempo, criou espaços virtuais e dispositivos tecnológicos baratos e fáceis de operar para a publicação diretamente na Grande Rede, expandindo o privilégio secular de editores e assemelhados. Um dos segmentos da sociedade que mais soube aproveitar esses espaços livres de publicação foi a comunidade acadêmica. No universo da pesquisa, as bibliotecas, paulatinamente, tornam-se os agentes que intermedeiam

esses espaços em termos de disponibilização de facilidades tecnológicas e de gestão. Para algumas modalidades de bibliotecas digitais – como é o caso dos repositórios institucionais e os repositórios temáticos – a auto-submissão, ou seja, a publicação diretamente na Rede pelo próprio autor é uma função vital. A arquitetura prevê um componente de serviço que disponibiliza, através de interfaces do portal do usuário, ferramentas para auto-submissão de autores que tenham autorização para tal. Os níveis de autorização, formato dos materiais, revisão dos conteúdos e opções de catalogação são definidos pelos gestores da biblioteca.

Temas de pesquisa

- Avaliação de usabilidade nos processos de auto-arquivamento.
- Padrões para auto-arquivamento.

Serviços de gestão / administração da biblioteca

Mesmo reconhecendo o lugar-comum da expressão administração de sistemas, no ambiente das bibliotecas digitais, seu significado ainda não está completamente definido. Um administrador da biblioteca digital controla o sistema em todos os aspectos: técnicos, gerenciais, legais, comerciais e éticos. Essa área pode não exigir pesquisas fundamentais, mas, decerto, demanda o desenvolvimento de ferramentas de administração mais orientadas para o problema. Uma questão-chave, por exemplo, é o estabelecimento de padrões para registro de atividades de bibliotecas digitais que permitam que diferentes sistemas possam ser comparados, que os hábitos de uso e consumo possam ser analisados e, ainda, que se possam estabelecer métricas para o desempenho dessas novas modalidades de biblioteca. Outra questão-chave é o aprofundamento da compreensão do conceito de qualidade de serviços para as bibliotecas digitais. (DIGITAL LIBRARIES, 2001).

No domínio da infra-estrutura proposta, os serviços de gestão e administração da biblioteca compreendem as funções e as ferramentas essenciais para o exercício das opções gerais de uma biblioteca digital particular, bem como para apoiar a busca pela conformidade com os aspectos legais, sociais e éticos pertinentes. Na grande área de administração de bibliotecas digitais, é necessário aprofundamento nos seguintes assuntos:

Temas de pesquisa

Inserção das bibliotecas digitais no comércio eletrônico de conteúdos.

Melhores práticas para bibliotecas digitais.

Métricas para avaliação de bibliotecas digitais.

Padrões para registro de atividades da biblioteca: sistema, uso e desempenho.

Planejamento de bibliotecas digitais: tecnologias, gestão e sustentabilidade.

Qualidade no contexto das bibliotecas digitais.

Serviços de gestão / administração da biblioteca: serviços de customização

Os serviços de customização compreendem as funções e as ferramentas do sistema que dão apoio à execução das opções técnicas da biblioteca, tais como política e instrumentos de indexação (esquemas de metadados, vocabulários controlados, esquemas de classificação, algoritmos de indexação, etc.), *lay-out* das interfaces, padrões de digitalização, organização e

apresentação das informações, política e padrões de interoperabilidade, entre outros. Esse componente de serviços pressupõe, também, a existência de ferramentas de *workflow* voltadas para o controle de entrada de dados e de outros processos gerenciais. As escolhas técnicas definidas pelo serviço vão caracterizar o funcionamento da biblioteca e, em paralelo com outros aspectos, conferir-lhe personalidade.

Temas de pesquisa

- Otimização de processos em bibliotecas digitais.
- Planejamento de bibliotecas digitais: diretrizes, padrões e tecnologias.
- Uso de *workflow* na gestão de bibliotecas digitais.

Serviços de gestão / administração da biblioteca: serviço de gestão de acesso

De uma forma geral, a idéia de biblioteca digital leva ao pressuposto de que todas suas informações estão disponíveis livremente e o acesso anônimo é plenamente permitido. Entretanto, no mundo real, é necessário prever que o acesso a alguns materiais pode ser restrito, e que autores e editores necessitarão saber como suas publicações estão sendo usadas e por quem, conforme Lagoze *et al.* (2002) chamam a atenção.

No ambiente do sistema proposto, que tem um universo amplo e heterogêneo de usuários com níveis de intervenção diversificados, quando um usuário solicita alguma operação à biblioteca – como acesso a determinada coleção ou a inserção de um item – sua requisição passa pelo processo chamado de gestão de acesso. Esse processo se desenrola basicamente em dois passos. O primeiro é a autenticação, que estabelece a identidade do usuário. O segundo é a autorização, que circunscreve as operações que são permitidas ao demandante.

No contexto da arquitetura, a gestão de acesso contribui para a personalização através das transações de autenticação e autorização. De fato, ela inclui um componente chamado servidor de perfis de usuários, o qual armazena atributos associados ao usuário que podem ser recuperados pelas interfaces do portal do usuário e por outros serviços para ajustes e personalização. O servidor pode, por exemplo, armazenar informações sobre o histórico do usuário, preferências, experiências, necessidades especiais; pode armazenar, ainda, dados demográficos e outras informações agregadas. No entanto, em qualquer circunstância, deve ser dada atenção às questões éticas e de privacidade.

Temas de pesquisa

- Anotação pessoal em objetos digitais.
- Bases de dados de perfis de usuários.
- Espaços comunitários informacionais.
- Gestão de acesso para bibliotecas digitais.

Serviços de gestão / administração da biblioteca: serviço de gestão de direitos

Compreende as funções gerenciais da biblioteca voltadas para a gestão de direitos autorais dos conteúdos digitais. Essa é uma questão bastante sensível, porque, na maior parte das vezes, a biblioteca é custodiante dos itens e das coleções digitais, mas não detém os direitos sobre eles, fato que tem impactos importantes na disseminação e na entrega ao usuário-final desses conteúdos.

Pesquisas relacionadas com o *copyright* recaem numa área nevrálgica, cercadas de inúmeros interesses e de conflitos permanentes. O ténue equilíbrio entre direito de acesso à informação e a recompensa econômica ao criador se torna dramaticamente instável com o surgimento da informação eletrônica distribuída via internet. Isso acontece porque no ambiente digital o objeto é menos fixado, o controle sobre as cópias é extremamente frágil e ele pode ser acessado remotamente por uma multiplicidade de usuários ao mesmo tempo.

Desde o ano 2000, essas pesquisas foram redimensionadas pelo crescente consenso da necessidade de proteção legal aos conteúdos disponibilizados por bibliotecas digitais, antagonizadas pelas discussões sobre direito pleno à informação e ao acesso livre. (LIEW, 2009). As doutrinas do *first sale* e o conceito do *fair use*, que formam as bases das bibliotecas acadêmicas tradicionais, não são facilmente transferidos para o universo das bibliotecas digitais e precisam ser reescritos, ou melhor, reinventados, considerando os novos e os futuros contextos.

Temas de pesquisa

- *Copyright* e preservação digital.
- *Digital Rights Management* (DRM) e bibliotecas digitais.
- *First sale* e *fair use* no acesso à informação digital.
- Propriedade intelectual no contexto das bibliotecas digitais.

Serviços de gestão / administração da biblioteca: serviço de preservação das coleções digitais

Se uma coleção em papel for deixada sem a devida manutenção, terá uma sobrevida inercial e as informações que guarda – dependendo das condições ambientais – poderão ser acessadas por longo tempo. O mesmo não acontece com as coleções digitais que necessitam de uma intencionalidade para que mantenham a capacidade de ser acessadas e ter suas informações interpretadas, posto que estão continuamente ameaçadas por dois inimigos vorazes: a obsolescência tecnológica e a degradação física das mídias. (SAYÃO, 2008).

A preservação digital consiste em atividade complexa e dispendiosa que se apóia num tripé cujos sustentáculos são a gestão, a tecnologia e o suporte institucional. Tomando por base esse tripé, observa-se que a decisão sobre a aderência às normas e aos padrões, a adoção das estratégias de preservação digital mais corretas e apropriadas, a aplicação de melhores práticas são responsabilidades gerenciais, e são viabilizadas por um conjunto de tecnologias. No entanto, todo processo depende, fortemente, de aportes financeiros e políticos da organização onde a biblioteca está inserida.

Nessa direção, a arquitetura proposta deve dispor de ferramentas técnicas e gerenciais que apoiem a preservação das coleções e de estruturas, principalmente de armazenamento, que permitam a compatibilidade com os padrões e os protocolos mais importantes, como o Modelo de Referência *Open Archival Information System* (OAIS), por exemplo.

Temas de pesquisa

- Aplicação do modelo de referência OAIS.
- Metodologias de gestão de preservação de acervos digitais.
- Modelos de custo para a preservação digital.
- Metadados para preservação digital.

- Repositórios confiáveis.
- Técnicas automatizadas para migração de conteúdos digitais.

Serviços de interoperabilidade

No domínio de pesquisa de bibliotecas digitais, interoperabilidade é, provavelmente, o problema mais intensamente discutido; a própria proposta de uma arquitetura aberta tem como pressuposto básico a interoperabilidade. A exigência por alcançar a interoperabilidade tem origem no fato de que várias bibliotecas digitais, diferentes em termos de arquitetura, de tecnologia e que utilizam diferentes esquemas de metadados e que são administradas por instituições distintas, desejam efetivamente interagir. Na perspectiva do usuário, essa interação deve acontecer de forma transparente e as informações integradas através de uma única interface. Fator de fundamental importância para a interoperabilidade são os padrões e os protocolos abertos que devem ser aplicados em todas as instâncias. No domínio bibliográfico, atualmente, dois protocolos, de enfoques distintos, exercem papel-chave, o OAI-PMH e o Z39.50. Entretanto, vários outros protocolos, de usos mais gerais, têm sido usados, com mais frequência, pelas bibliotecas digitais para alcançar a interoperabilidade, o *Web Service* é um exemplo significativo desse fato.

A interoperabilidade se completa por várias instâncias. A mais óbvia delas é a interoperabilidade técnica, dependente de tecnologias e protocolos. Por outro lado, a interoperabilidade semântica necessita de instrumentos comuns de representação de conhecimento; e a interoperabilidade política exige acordos e convenções entre as organizações.

Temas de pesquisa

- Agentes e ontologias na federação de bibliotecas digitais.
- Aplicação de tecnologias de Web semântica em bibliotecas digitais.
- Bibliotecas digitais baseadas em ontologias.
- Federação de bibliotecas digitais heterogêneas.
- Interoperabilidade baseada em *Web Service*.
- Mapeamento entre esquemas de metadados.

Ampliando o escopo de pesquisa em bibliotecas digitais

O discurso das bibliotecas digitais tem sido veemente não somente nos domínios óbvios da tecnologia, dos *e-services* e da arquitetura de informação. Muitas outras disciplinas fora do escopo da Biblioteconomia e dos estudos da CI compõem a matriz de conhecimentos necessários para dar conta dos desafios da área de bibliotecas digitais. A abrangência interdisciplinar inclui itens da área de administração, tais como finanças, *marketing* e logística, bem como itens concernentes à educação, à política e à regulamentação e, ainda, aspectos sociais e culturais. Também é impossível deixar de lado questões éticas, legais e a dimensão de uso. Estes temas ganham relevância na medida em que as bibliotecas digitais se tornam ambientes de conhecimento e aprendizado. Liew (2009) apresenta um mapeamento sobre as pesquisas em torno dos aspectos organizacionais e humanos concernentes às bibliotecas digitais que indica alguns temas importantes de pesquisa.

Temas de pesquisa

- Algoritmos e técnicas de tradução aplicadas às bibliotecas digitais.
- Aspectos éticos das bibliotecas digitais: privacidade, confidencialidade, confiabilidade.
- Aspectos econômicos e organizacionais das bibliotecas digitais: modelos de sustentabilidade e financiamento, cooperativismo e planejamento.
- Aplicações: educação, medicina, entretenimento, Governo, meio-ambiente.
- Biblioteca digital como instrumento de inclusão e de emponderamento social.
- *Marketing*, promoção e divulgação de bibliotecas digitais.
- Modelo de bibliotecas digitais multilíngües.

À guisa de conclusão

A idéia de problematizar, no sentido de trazer à tona uma abordagem contextualizada e interdisciplinar, os temas de pesquisa para a área de bibliotecas digitais, impõe desafios metodológicos não enfrentados nessa primeira abordagem. O que se propôs foi mero exercício – que pode sempre ser retomado com maior rigor e precisão – que demonstrasse a necessidade de que qualquer análise da área de pesquisa e de aplicações em bibliotecas digitais requer, antes de tudo, compreensão clara de todos os atores presentes e de seus relacionamentos. Essas variáveis, por terem muitas faces, precisam ser modeladas para ser vistas de seus ângulos mais corretos e mais favoráveis. A arquitetura conceitual, dessa forma, foi apenas o retrato que serviu como fonte para identificação de problemas e lacunas que poderão ser objeto de aprofundamento em termos acadêmicos ou de prática profissional. O modelo serve, ainda, para demonstrar a necessidade de arquiteturas mais ousadas que sirvam de infra-estrutura para a construção de novas concepções de bibliotecas digitais.

O que se verifica, com certa clareza, é que a amplitude dos estudos e das pesquisas na área de bibliotecas digitais compreende parte considerável do domínio de aplicações da Biblioteconomia e da CI. Por outro lado, o campo de bibliotecas digitais tem interesse na totalidade do desenvolvimento da área de tecnologia da informação, sobretudo, quando se trata de aplicações em informação e conhecimento. Todas essas aplicações estão mapeadas no vasto campo de pesquisa das bibliotecas, dos museus e dos arquivos digitais.

Por fim, é necessário dizer que a imensa convergência de conhecimentos que representa a pesquisa sobre bibliotecas digitais reflete o desejo comum de melhor compreender a natureza dos sistemas distribuídos de informação, do acesso ao conhecimento num ambiente digital e de seus impactos. Por conta disso, espera-se, num futuro previsível, que pesquisas inovadoras sobre essas bibliotecas, em termos de tecnologias e de aplicações, contribuam significativamente para o incremento da utilização de informações eletrônicas de todo tipo, e que seus benefícios se espalhem por toda sociedade sem distinções e barreiras.

REFERÊNCIAS

AGER, Tryg. Architecture and system. In: INTERNATIONAL TECHNOLOGY RESEARCH INSTITUTE. **Digital information organization in Japan**. 1999. Baltimore: DARPA, 1999. p. 23- 40. Disponível em: <<http://www.wtec.org/loyola/pdf/dio.pdf>>. Acesso em: 3 jul. 2009.

ARMS, W. **Digital libraries**. Cambridge, Mass.: MIT Press, 2000. Disponível em: <<http://www.cs.cornell.edu/wya/diglib/MS1999/index.html>>. Acesso em: 9 jul. 2009.

CANDELA, Leonardo *et al.* A reference architecture for digital library systems. **Ercim News**, [S. l.], n. 66, Jul. 1966. Disponível em: <http://www.ercim.org/publication/Ercim_News/enw66/candela.html>. Acesso em: 9 jul. 2009.

CARIGNAN, Yvonne *et al.* **Best practice guidelines for digital collection**. Maryland: University of Maryland, 2007. 76 p. Disponível em: <http://www.lib.umd.edu/dcr/publications/best_practice.pdf>. Acesso em: 6 jul. 2009.

CÉSPEDES, Zulia. Critérios e indicadores para evaluar las bibliotecas digitales. **Acimed**, [S. l.], v. 14, n. 6, 2006.

DIGITAL LIBRARIES: future directions for a European Research Programme. Brainstorm Report, 2001. Disponível em: <<http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf>>. Acesso em: 2 jul. 2009.

JANÉE, Greg; FREW, James. The ADEPT digital library architecture. In: INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES, 2., 2002, Portland. **Proceedings...** New York: ACM, 2002. p. 342-350. Disponível em: <<http://portal.acm.org/citation.cfm?id=544220.544264>>. Acesso em: 2 jul. 2009.

KAHN, Robert; WILENSKY, Robert. A framework for distributed digital object services. **International Journal on Digital Libraries**, [S. l.], v. 6, n. 2, 2006. Disponível em: <<http://www.springerlink.com/content/0723r55h83067n10/fulltext.pdf>>. Acesso em: 9 jul. 2009.

LAGOZE, Carl *et al.* Core services in the architecture of the National Science Digital Library (NSDL). In: INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES, 2., 2002, Portland. **Proceedings...** New York: ACM, 2002. p. 201-209. Disponível em: <<http://portal.acm.org/citation.cfm?id=544220.544264>>. Acesso em: 2 jul. 2009.

LAGOZE, Carl; PAYETTE, Sandra. **An infrastructure for open-architecture digital libraries**. Cornell University, 1998. Disponível em: <<http://ecommons.library.cornell.edu/bitstream/1813/7344/1/98-1690.pdf>>. Acesso em: 2 jul. 2009.

LIEW, Chern Li. Digital library research 1997-2007: organizational and people issue. **Journal of Documentation**, [S. l.], v. 65, n.2, p. 245-266, 2009.

NATIONAL INFORMATION STANDARDS ORGANIZATION. NISO FRAMEWORK WORK GROUP. **A framework for guidance for building good digital collection**. Baltimore, Dec. 2007. Disponível em: <<http://www.niso.org/publications/rp/framework3.pdf>>. Acesso em: 3 jul. 2007.

PANDEY, Richa. **Digital library architecture**. In: DRTC WORKSHOP ON DIGITAL LIBRARIES: THEORY AND PRACTICE, Bangalore, 2003. p. 1-16. Disponível em: <https://drtc.isibang.ac.in/bitstream/handle/1849/22/B_architecture_richa.pdf?sequence=2>. Acesso em: 8 jul. 2009.

PAYETTE, Sandra. **Digital library architecture: a service-based approach**. 1998. Disponível em: <<http://www2.cs.cornell.edu/payette/presentations/DL-architecture.ppt>>. Acesso em: 2 jul. 2009.

POWEL, Andy. **Notes about possible technical criteria for evaluating institutional repository (IR) software**. Bath: University of Bath, Dec. 2005. Disponível em: <<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/ir-software.pdf>>. Acesso em: 2 jul. 2009.

POWEL, Andy; LYON, Liz. **The DNER technical architecture: scoping the information environment**. Bath: University of Bath, May 2001. Disponível em: <<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html>>. Acesso em: 3 jul. 2009.

SAYÃO, Luís Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes – URN, PURL, DOI, *Handle System*, *CrossRef* e *OpenURL*. **Transinformação**, Campinas, v. 19, n. 1, 2007.

SAYÃO, Luís Fernando. Preservação de revistas eletrônicas. In: FERREIRA, S. M. S. P; TARGINO, M. das G. (Org.). **Mais sobre revistas eletrônicas: em foco a gestão**. São Paulo: SENAC, 2008. 221 p.

SCHAUBLE, Peter; SMEATON, Alan F. (Ed.) **An international research agenda for digital libraries: summary report of the series of joint NSF-EU working groups on future directions for digital libraries research**. Oct. 12, 1998. Disponível em: <http://www.ercim.org/publication/ws-proceedings/DELOS-B/dl_sum_report.pdf>. Acesso em: 2 jul. 2009

SHIRI, Ali. Digital library research: current developments and trends. **Library Review**, [S. l.], v. 52, n. 5, p. 198-202, 2003. Disponível em: <http://www.cis.strath.ac.uk/cis/research/publications/papers/strath_cis_publication_39.pdf>. Acesso em: 2 jul. 2009.

SULEMAN, Hussein; FOX, Eduard. A framework for building open digital libraries. **D-Lib Magazine**, [S. l.], v.7, n.12, 2001. Disponível em: <<http://www.dlib.org/dlib/december01/suleman/12suleman.html>>. Acesso em: 4 jul. 2009.

BIBLIOGRAFIA

CASTELLI, Donatella; PAGANO, Pasquale; THANOS, Constantino. OpenDLib: na infrastructure for new generation digital libraries. **International Journal on Digital Libraries**, [S. l.], v. 4, n. 1, p. 45-47, Aug. 2004.

LAGOZE, Carl; FIELDING, David. Defining collection in distributed digital libraries. **D-Lib Magazine**, [S. l.], Nov. 1998. Disponível em: <<http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>>. Acesso em: 2 jul. 2009.

SAYÃO, Luis Fernando; MARCONDES, Carlos H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 02, 2008.