



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL

LUIZ PAULO MARTINS FIGUEIREDO

**IMPLEMENTAÇÃO DE APLICAÇÃO PARA CONSUMIR
SERVIÇO WEB REFERENTE À PRESERVAÇÃO
DIGITAL**

Bandeirantes

2012

LUIZ PAULO MARTINS FIGUEIREDO

**IMPLEMENTAÇÃO DE APLICAÇÃO PARA CONSUMIR
SERVIÇO WEB REFERENTE À PRESERVAÇÃO
DIGITAL**

Monografia apresentada à Universidade Estadual do Norte do Paraná – *campus* Luiz Meneghel – como requisito parcial para a obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof.^a Cristiane Yanase Hirabara Castro

Bandeirantes

2012

LUIZ PAULO MARTINS FIGUEIREDO

**IMPLEMENTAÇÃO DE APLICAÇÃO PARA CONSUMIR
SERVIÇO WEB REFERENTE À PRESERVAÇÃO
DIGITAL**

Monografia apresentada à Universidade Estadual do Norte do Paraná – *campus* Luiz Meneghel – como requisito parcial para a obtenção do grau de Bacharel em Sistemas de Informação.

COMISSÃO EXAMINADORA

Prof. Cristiane Yanase Hirabara Castro
UENP – *Campus* Luiz Meneghel

Prof. Estevan Braz Brandt Costa
UENP – *Campus* Luiz Meneghel

Prof. José Reinaldo Merlin
UENP – *Campus* Luiz Meneghel

Bandeirantes, __ de _____ de 2012

*Onde não falta vontade
existe sempre um
caminho.
(O Senhor dos Anéis)*

RESUMO

Este trabalho refere-se a uma aplicação que consome serviços web disponíveis na internet e que podem ser utilizados para preservação digital. Como a tecnologia avança cada vez mais rápido, precisamos de técnicas que ajudem a preservar os documentos por mais tempo mantendo assim suas propriedades com o passar do tempo. A aplicação se resume a combinar em um único lugar alguns serviços web de preservação digital.

Palavras-chave: Serviços Web. Biblioteca Digital. Preservação Digital.

ABSTRACT

This work refers to an application that consumes web services available on the Internet and can be used for digital preservation. As technology advances faster than ever, we need techniques that help to preserve the documents for a longer time thus maintaining their properties over time. The application comes down to combine in one place some web services for digital preservation.

LISTA DE SIGLAS

BEPRESS	<i>Berkeley Electronic Press</i>
CD	<i>Compact Disc</i> (Disco Compacto)
CRiB	<i>Conversion and Recommendation of Digital Object Formats</i>
DAITSS	<i>Dark Archive In The Sunshine State</i>
DVD	<i>Digital Versatile Disc</i> (Disco Digital Versátil)
Fedora	<i>Flexible Extensible Digital Object Repository Architecture</i> (Arquitetura de Repositório Extensiva Flexível para Objetos Digitais)
HTTP	<i>Hypertext Transfer Protocol</i> (Protocolo de transferência de hipertexto)
JHOVE	<i>JSTOR/Harvard Object Validation Environment</i>
OAIS	<i>Open Archival Information System</i> (Sistema Aberto para Arquivamento de Informação)
OASIS	<i>Organization for the Advancement of Structured Information Standards</i> (Organização para o Avanço de Padrões em Informação Estruturada)
PD	Preservação Digital
PUID	<i>PRONOM Unique IDentifier</i> (Identificador Único do PRONOM)
SOA	<i>Service Oriented Architecture</i> (Arquitetura Orientada a Serviço)
SOAP	<i>Simple Object Access Protocol</i> (Protocolo Simples de Objeto de Acesso)
UDDI	<i>Universal Description Discovery and Integration</i> (Descrição Universal de Descoberta e Integração)
UML	<i>Unified Modeling Language</i> (Linguagem Unificada de Modelagem)
WS	<i>Web Services</i> (Serviço Web)
WSDL	<i>Web Service Definition Language</i> (Definição de Linguagem de Serviços Web)
XML	<i>eXtensible Markup Language</i> (Linguagem de Marcação Estendida)

LISTA DE FIGURAS

Figura 1 - Comunicação do cliente com o servidor	16
Figura 2 - Estrutura de uma mensagem SOAP	17
Figura 3 - Comunicação entre Cliente, Provedor de Serviço e Documento WSDL.....	18
Figura 4 - Usuário e Provedor utilizando UDDI.....	19
Figura 5 - Evolução das Mídias	22
Figura 6 - Modelo de Referência Open Archival Information System (OAIS).....	25
Figura 7 - Esquema de Identificação de Assinatura Externa e Interna.....	32
Figura 8 - Arquitura CRiB	34
Figura 9 - Utilização do JHOVE.....	35
Figura 10 - Tela de Verificação	38
Figura 11 - Tela de Seleção do Arquivo.....	39
Figura 12 - Arquivo não selecionado.....	40
Figura 13 - Tela de Espera.....	41
Figura 14 - Tela de Verificação Concluída	42
Figura 15 - Diagrama de Caso de Uso.....	43
Figura 16 - Diagrama Classe Análise.....	45
Figura 17 - Diagrama de Comunicação	46
Figura 18 - Diagrama de Sequência.....	47
Figura 19 - Diagrama de Estados Navegacionais	48
Figura 20 - Diagrama de Atividade	49

SUMÁRIO

1	INTRODUÇÃO.....	11
1.1	FORMULAÇÃO DO PROBLEMA	12
1.2	OBJETIVOS	12
1.2.1	Objetivo Geral.....	12
1.2.2	Objetivos Específicos.....	12
1.3	JUSTIFICATIVA.....	13
1.4	Metodologia	13
2	SERVIÇOS WEB	14
2.1	Arquitetura	15
2.2	<i>Simple Object Access Protocol (SOAP)</i>	15
2.3	<i>Web Service Description Language (WSDL)</i>	17
2.4	<i>Universal Description, Discovery and Integration (UDDI)</i>	18
3	PRESERVAÇÃO DIGITAL.....	20
3.1	Estratégias de Preservação Digital	22
3.1.1	Preservação da Tecnologia	22
3.1.2	Migração ou Conservação	23
3.1.3	Encapsulamento	23
3.1.4	Refrescamento	24
3.1.5	Conversão para Formatos Concorrentes.....	24
3.2	Modelo Conceitual <i>Open Archival Information System (OAIS)</i>	24
3.3	Repositórios Digitais	26
3.3.1	DSpace	27
3.3.2	<i>Flexible Extensible Digital Object Repository Architecture (Fedora)</i>	27
3.4	Bibliotecas Digitais.....	27
3.5	Serviços de Preservação Digital	30
3.5.1	Serviço Provido Pelo PRONOM	30
3.5.2	<i>Conversion and Recommendation of Digital Object Formats (CRiB)</i>	33
3.5.3	<i>JSTOR/Harvard Object Validation Environment (JHOVE)</i>	35
4	PROJETO.....	36
4.1	Material	36

4.1.1	Linguagem de programação	36
4.1.2	API's e Bibliotecas	36
4.1.3	Ambiente de Desenvolvimento	37
4.2	APWS - Aplicação de Serviços Web	37
4.2.1	Modelagem da Arquitetura do Sistema.....	42
5	DISCUSSÃO DE RESULTADOS	51
6	CONCLUSÃO	52
	REFERÊNCIAS.....	53

1 INTRODUÇÃO

Estamos vivendo na era da informação, e como o próprio nome diz a informação é um dos fatores essenciais desse período. Com os preços mais baixos dos dispositivos de armazenamento guardar uma grande quantidade de informação ficou mais fácil que em outros tempos não tão distantes.

Uma das grandes interessadas na questão da informação são as empresas, pois elas podem ter uma visão melhor de como estão os negócios e realizar estimativas para uma futura ação. Uma das grandes revoluções da tecnologia foi à troca de informação instantaneamente, nunca foi tão fácil trocar informações com uma pessoa que esta do outro lado do mundo e de forma tão rápida. A importância da informação na internet cresceu muito nesses últimos anos, antes a informação que estava restrita apenas as bibliotecas, agora pode ser encontrada facilmente na internet.

O que se tem discutido bastante esses últimos tempos em computação é a Preservação Digital (PD), pois com as rápidas evoluções tecnológicas deve-se desenvolver métodos para que os documentos tenham uma longa duração. O objetivo da PD é criar estratégias e métodos para manter os documentos ao longo dos anos, sem que os documentos percam seu conteúdo intelectual.

Atualmente é fácil identificar a evolução tecnológica das mídias, por exemplo, primeiramente era utilizado o disquete, depois foi desenvolvido o CD e logo após o DVD, e com esse desenvolvimento tecnológico muitas vezes os documentos são perdidos por as técnicas para a recuperação estarem obsoletas.

Nas universidades os estudantes sempre necessitam de materiais que possam ser utilizados como fonte de pesquisa, e a internet é o local mais procurado por vários estudantes por causa da facilidade de acesso e o grande número de informações. Encontram-se na maioria documentos que não trazem tanta confiabilidade e muitas vezes que trazem informações corretas, mas que não são, por exemplo, uma publicação oficial como um artigo ou monografia.

1.1 FORMULAÇÃO DO PROBLEMA

Como a tecnologia esta evoluindo cada vez mais rápido, é preocupante o fato, em que alguns documentos de grande interesse as pessoas acabem entrando em obsolescência. Livros, artigos, monografias, entre outros documentos precisam ser preservados e a preservação digital é o que se tem discutido bastante. Criar formas para que os documentos acessados hoje possam ser utilizados futuramente da mesma maneira, mantendo suas propriedades e características. O processo de Preservação Digital mantém os objetos digitais em repositórios, mas antes que isso aconteça é preciso realizar uma identificação do objeto em si, para poder armazenar todas as suas características e que posteriormente ele possa ser acessado de forma satisfatória.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo deste trabalho é criar uma interface que utilize um serviço que possa auxiliar na preservação digital, isto é, uma interface para consumação de serviço web disponibilizado por arquiteturas existentes e consolidadas.

1.2.2 Objetivos Específicos

Os objetivos específicos são:

- Selecionar algum serviço web para preservação digital disponível na web.
- Modelar a interface
- Propor uma arquitetura para interface de serviço web de preservação digital
- Desenvolver uma aplicação que consuma o serviço web.
- Testar o serviço web que foi utilizado.

1.3 JUSTIFICATIVA

Com o passar do tempo os documentos vão ficando velhos e muitas vezes desaparecem, o que se apresenta necessário é alguma forma que faça esses documentos perdurarem pelos anos que se seguirão. Uma alternativa para a longa duração desses documentos é a preservação digital, que mantém documentos que já se encontram digitalmente em um repositório e passa os documentos físicos para um formato digital para que se faça o mesmo processo. Para que futuramente esses documentos ainda possam ser acessados é necessário que se crie padrões de preservação digital, por isso quando falamos em preservação digital hoje em dia uma das primeiras tecnologias que aparecem é o *eXtensible Markup Language* (XML), que parece ser o mais indicado para que esses documentos durem bastante tempo, pois independe de linguagem de programação.

O objetivo deste trabalho é o desenvolvimento de uma aplicação que absorva serviços web de preservação digital que estejam disponíveis na internet, para isso há uma seleção de alguns serviços web em um único lugar para que os documentos passem por uma padronização antes de serem dirigidos aos repositórios.

1.4 Metodologia

Dando enfoque aos métodos, divide-se em duas partes: pesquisa exploratória, pois tem uma fundamentação em materiais que já foram publicados anteriormente, e a implementação, desenvolvimento de uma aplicação para ratificar a solução proposta.

A respeito da aplicação realizou-se uma pesquisa para melhor compreensão dos aspectos referentes a serviços web e sobre os serviços que provavelmente poderiam ser utilizados no desenvolvimento.

2 SERVIÇOS WEB

Todas as principais plataformas têm acesso a internet através de navegadores da Web, plataformas diferentes normalmente não poderiam interagir, mas para que elas trabalhem em conjunto foram desenvolvidos *Web Services*. São aplicativos que funcionam através da internet, construídos com padrões de navegadores da Web garantindo assim que seja usado por qualquer navegador em qualquer plataforma (W3SCHOOLS Acesso em: 27 de junho de 2012).

Segundo Basiura (2003) um *Web Service* é uma lógica de aplicação acessível a programas protocolos-padrão presentes na Web de maneira que independe da plataforma utilizada. Um *Web Service* é acessado através de um programa que pode ser uma aplicação web, uma aplicação desktop, ou qualquer uma que consiga realizar uma conexão com o serviço (BASIURA, 2003).

Na internet circulam várias informações importantes, e as empresas tem que garantir que seus dados sejam seguros, com os serviços web as aplicações tornam-se mais vulneráveis expondo assim fluxos de negócios, processos e arquitetura. Mecanismos estão sendo propostos em serviços web, mas ainda não atendem a todas as necessidades de segurança exigidas (MELLO, 2006).

Segundo Cruz (2005) uma das principais vantagens que os serviços web trazem é a interoperabilidade obtida através da adesão de padrões que são amplamente utilizados na web, como Protocolo Simples de Acesso a Objetos (SOAP), Protocolo de Transferência de Hipertexto (HTTP), Linguagem de Marcação Extensível (XML), e outros mais.

Os serviços web cada vez mais ganham seu espaço, pois estão presentes em diversos mercados. Eles têm como função a construção de redes intra e interorganizacionais de aplicações distribuídas e colaborativas, onde são invocados através de uma arquitetura orientada (CRUZ, 2005).

Os *Web Services* podem ser vistos como um processo de terceirização, onde a aplicação que o consome se equipara a empresa principal e o serviço web propriamente dito seriam as empresas terceirizadas que prestam serviços a principal.

2.1 Arquitetura

Os serviços web podem ser definidos como uma arquitetura baseada em serviços, que é mais conhecida por *Service Oriented Architecture (SOA)*, termo primeiramente utilizado pela IBM. A arquitetura dos serviços web representa uma junção de serviços que se comunicam através de troca de mensagens XML. Essa arquitetura possui três tipos de papéis que comunicam entre si:

- Provedor de serviço – descreve e publica um determinado serviço Web no registro dos serviços. Descreve também as informações de ligação do serviço que são utilizadas para sua chamada. As informações utilizadas nessa arquitetura são representadas num documento XML escritas em linguagem padrão *Web Service Description Language (WSDL)*;
- Consumidor de serviço – responsável por encontrar um serviço, obter a descrição dele e utilizá-lo para fazer a comunicação entre ele e um provedor e assim invocar um serviço Web;
- Registro dos Serviços – guarda informações a respeito dos serviços. O padrão que o SOA utiliza para o registro de informações é o *Universal Description Discovery and Integration (UDDI)* (CRUZ, 2005).

O SOA pode ser dito como uma abordagem de arquitetura para o desenvolvimento de sistemas criados a partir de serviços autônomos. A integração nesse tipo de arquitetura é o que se tem dado bastante ênfase, pois provavelmente uma solução final será composta de diversos serviços feitos em linguagens diferentes, hospedados em plataformas diferentes. A real função dessa arquitetura seria a independência de linguagens e plataformas, seriam serviços disponibilizados que poderiam ser acessados pela maioria (EVDEMON, 2005).

2.2 *Simple Object Access Protocol (SOAP)*

O protocolo SOAP é um protocolo de comunicação simples que permite a troca de informações entre aplicações usando o protocolo HTTP, é um protocolo utilizado para acessar serviços Web. Permite a comunicação entre programas utilizando-se da internet como meio. Como o HTTP é a melhor maneira para as aplicações se

comunicarem e é suportado por todos os navegadores da internet, o SOAP foi desenvolvido para melhorar essa comunicação trabalhando com XML que é uma linguagem padrão para internet (CANÊS, 2011).

A comunicação que acontece entre o cliente utilizando o protocolo SOAP e o servidor que disponibilizando o serviço web, fazendo as requisições e obtendo as respostas pode ser vista na Figura a seguir:

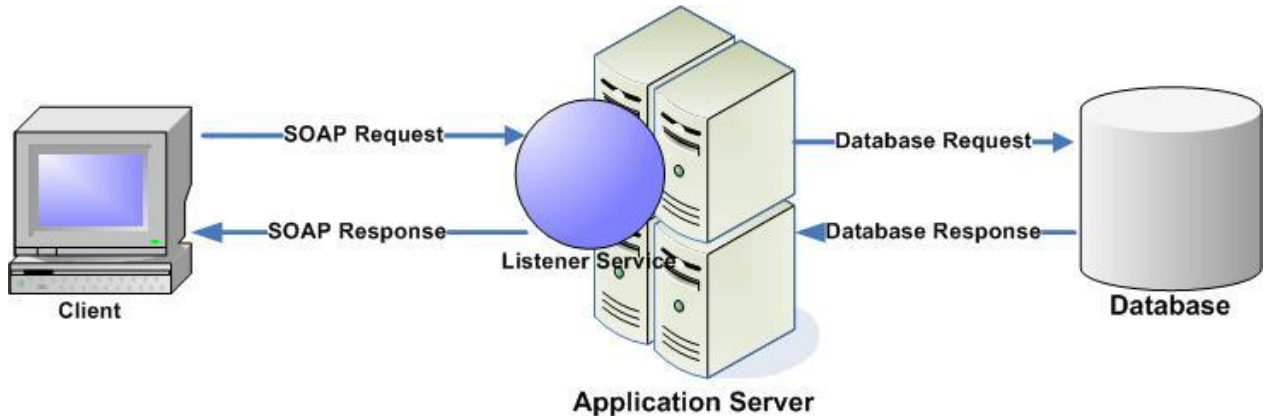


Figura 1 - Comunicação do cliente com o servidor

Fonte : <http://community.citrix.com/>

Uma mensagem SOAP é um documento XML que contém os seguintes elementos:

Envelope: Toda mensagem SOAP deve conter esse elemento, é um dos principais do documento XML, ele pode conter declarações de *namespaces* e atributos adicionais.

Header: Cabeçalho adicional que é utilizado para carregar informações adicionais, quando utilizado deve ser o primeiro a aparecer no Envelope.

Body: Elemento obrigatório que contém o *payload*, ou uma informação que é transportada ao destino final. Dentro dele também existe o elemento *Fault*.

Fault: Carrega e processa mensagens de erros que são enviadas pelos nós ao processarem a mensagem (DANTAS, 2007).

A estrutura do envelope do protocolo SOAP pode ser visto na Figura 2:

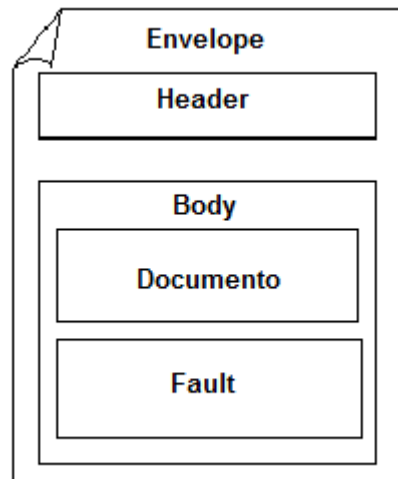


Figura 2 - Estrutura de uma mensagem SOAP.
Fonte: Elaborado pelo Autor

2.3 Web Service Description Language (WSDL)

Web Service Description Language (Linguagem de Descrição de Serviço Web) é uma linguagem utilizada para descrever *Web Service* ou pontos finais de rede baseada em XML. Um contrato que realiza a conexão entre o serviço web e o cliente descrevendo as mensagens, a localização do serviço e os protocolos que estão disponíveis para realizar a comunicação. O WSDL funciona em conjunto com o SOAP e a UDDI para que os serviços web possam interagir com outros *Web Services*, aplicações e dispositivos que acessam através da internet (BASIURA, 2003).

O WSDL é um documento em XML utilizado para descrever serviços web, ele faz com que não existam detalhes concretos sobre a descrição do serviço, a pessoa que utiliza esse serviço não tem informações de onde o serviço está, ou como ele é oferecido. É uma especificação que define a funcionalidade abstrata de um serviço (CHINNICI, 2002).

A relação estabelecida entre o provedor de serviço, o cliente e o documento WSDL pode ser vista na Figura 3 a seguir:

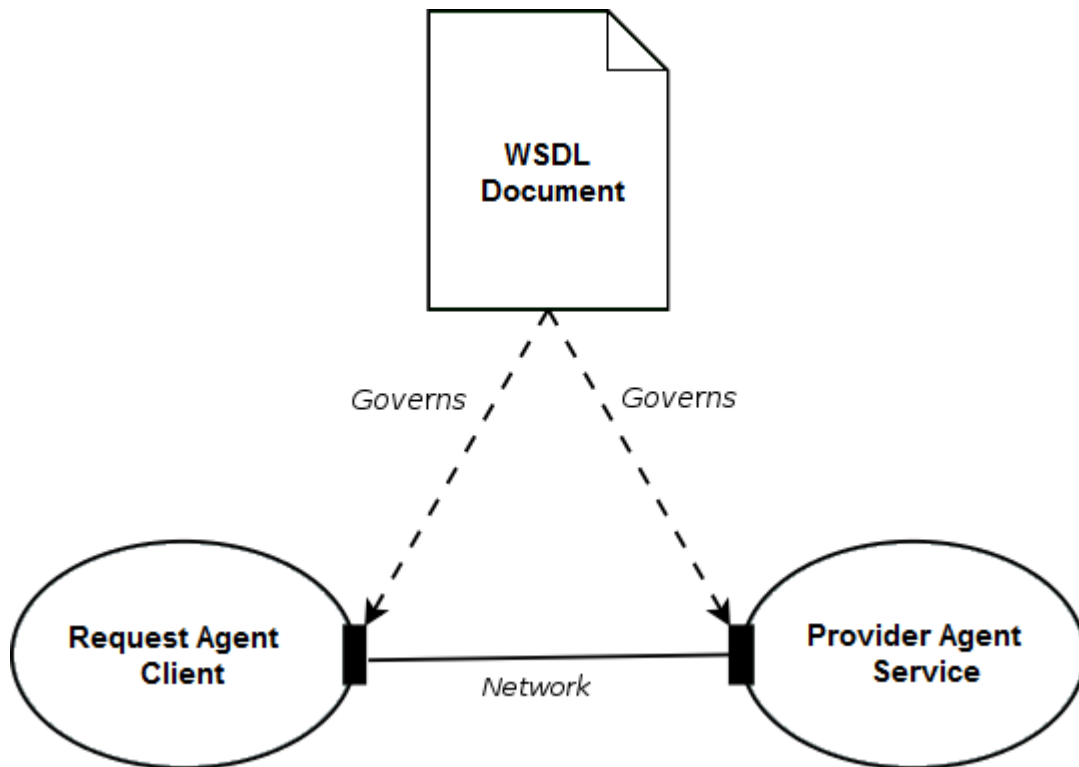


Figura 3 - Comunicação entre Cliente, Provedor de Serviço e Documento WSDL
 Fonte: <http://www.w3.org/>

2.4 Universal Description, Discovery and Integration (UDDI)

UDDI oferece um banco de dados que independe de plataforma para a identificação da localização e categoria das empresas e *Web Services*. Ele funciona como uma ferramenta para procurar e localizar organizações em nossa área de atuação ou aquela em que estamos interessados. Os repositórios UDDI independem de plataforma fazendo assim que qualquer organização externa ou individuo tenha a possibilidade de inserir informações no UDDI (BASIURA, 2003).

O UDDI é um protocolo padrão aprovado pela *Organization for the Advancement of Structured Information Standards* (OASIS), ele é de fundamental importância na pilha de serviços que a web oferece. Método padrão para a publicação e descoberta de componentes de rede baseados em *softwares* que utilizam a arquitetura orientada a serviços (SOA) (UDDI.XML.ORG Acesso em: 27 de junho de 2012).

Na Figura 4 a seguir acontece a relação do usuário e o do provedor de serviço com o repositório UDDI, o primeiro busca informações referentes a organização que

esta oferecendo o serviço e o segundo gravando informações referentes a seu serviço para que possa ser encontrada na utilização do mesmo.



Figura 4 - Usuário e Provedor utilizando UDDI
Fonte: <http://www.nuwanbando.com>

3 PRESERVAÇÃO DIGITAL

As diversas formas de armazenamento digital são incertas quanto a sua duração, pois possuem um tempo que é relativamente curto para que possa atingir as novas gerações. (SÁ, 2009)

Segundo Keefer (2005) diferentemente dos métodos de preservação tradicionais como os utilizados pelos museus e bibliotecas, ou seja, as instituições que mantêm a história, as aplicações digitais não conseguem seguir o mesmo tipo de armazenamento. Alguns dos perigos que esse tipo de armazenamento encontra é a dependência do ambiente de programação, facilidade de modificação dos dados (dificuldade de garantir a autenticidade e integridade dos documentos), entre outros.

De acordo com Lee (2002) o ambiente em que os equipamentos, *softwares* e formatos se encontram mudam constantemente por causa das evoluções tecnológicas e a preservação digital tem que garantir que haja acesso contínuo aos conteúdos dos documentos digitais. Não deve somente obter o objeto informacional, mas também seu significado, por isso deve se criar uma forma original que garanta acessibilidade e autenticidade.

A importância da preservação digital surgiu da necessidade que alguns profissionais tiveram de assegurar que os dados que foram gerados ou transformados para o formato digital não entrem em obsolescência. Se não acontecer uma resolução em relação a isso, a humanidade pode perder grande parte de informações que estão em suportes eletrônicos, pois eles podem ficar ultrapassados, ou em péssimas condições que o acesso à informação seja impossível ou muito difícil (JESUS, 2011).

Na maioria as bibliotecas não possuem uma política que seja para a preservação de seus documentos digitais ao longo do tempo sendo que muitas dessas informações não possam mais ser acessadas pelos usuários (SÁ, 2009).

Segundo Tammaro (2008) o que acontece é que muitos documentos não podem ser acessados nem utilizados sem as tecnologias apropriadas, esse problema acontece tanto nos documentos que são originalmente digitais quanto aqueles que foram transformados do analógico para o digital.

A escolha de uma estratégia efetiva de preservação é apenas o começo, pois ainda deve se garantir acesso aos documentos digitais e fazer com que eles não fiquem

obsoletos com as rápidas mudanças tecnológicas. O documento digital pode ser acessado através de determinado programa que se torna obsoleto com o tempo, existem variadas formas e modelos de preservação digital que devem ser examinados para poder decidir qual deles se enquadra melhor para o documento digital e que possa ser transferido mais fácil para as tecnologias que se seguirão. (PHALET, 2004)

Os documentos digitais ao passar do tempo se tornam inacessíveis por não ter mais programas que possam acessá-los por estes terem se tornados obsoletos. O que deve se fazer quanto a não obsolescência desses documentos é a definição de variadas medidas estratégicas e procedimentos para que eles possam sobreviver e manter sua autenticidade ao longo do tempo para que ocorra uma perda mínima do conteúdo informativo, acessibilidade e sua funcionalidade (TAMMARO, 2008).

Segundo Rothenberg (1998) a vida útil dos meios de armazenamento físicos algumas vezes é assombrosamente curta, por isso são necessárias informações a respeito dos objetos digitais para que se possa fazer um “refrescamento” transferindo-os para novas mídias. A obsolescência tecnológica desses meios de comunicação (*hardware e software*) é uma ameaça diferente e igualmente urgente. Os problemas referentes a preservação não estão somente no processo propriamente dito, mas também nas questões judiciais, financeiras, entre outras.

A Figura 5 mostra a evolução das mídias, do rolo de filme utilizado nos cinemas até o cartão de memória que pode armazenar uma grande quantidade de informação em um material relativamente pequeno:



Figura 5 - Evolução das Mídias
 Fonte: Elaborada pelo Autor

3.1 Estratégias de Preservação Digital

Com o passar do tempo tem se desenvolvido diversas técnicas para solucionar o problema da obsolescência tecnológica e com isso, vários autores dividiram de inúmeras formas diferentes, mas nesse trabalho será dividida em cinco: preservação da tecnologia, migração ou conversão, encapsulamento, refrescamento e conversão para formatos concorrentes.

3.1.1 Preservação da Tecnologia

Nesse tipo de estratégia o objetivo principal é manter o objeto como ele realmente é da maneira em que ele foi criado para que ele esteja na sua forma mais original. Os que impulsionaram essa tecnologia acreditam que seja uma das melhores, pois podem assegurar que o objeto seja visto da mesma forma que seria visto na data de sua criação (FERREIRA, 2007).

Esse tipo de estratégia pode ser entendida como manter o presente para que possa ser usado no futuro. Como vimos no decorrer dos anos a tecnologia avança muito rápido e é mais provável que se utilizem outros tipos de estratégias para preservação do que a apresentada.

3.1.2 Migração ou Conservação

Migração ou conservação pode ser entendida como a transferência do objeto de uma tecnologia em que ele está situado tanto *hardware* como *software* para uma posterior e realizar esse procedimento periodicamente. O objeto digital tem sua estrutura de elementos organizados de determinada forma, o que acontece na migração é a reorganização desses elementos para que eles possam constituir um objeto com uma nova estrutura e assim fazer parte de um novo tipo de tecnologia (FERREIRA, 2007).

A migração periódica da informação digital seja de *hardware* ou *software* para outro é a estratégia de PD mais comumente utilizada pelas instituições que possuem grandes acervos (ARELLANO, 2004).

A migração nada mais é do que fazer com que os objetos digitais sigam evoluindo junto com as tecnologias, pode ser comparado atualmente com a transferência de mídias em cassete para as mídias digitais como o DVD.

3.1.3 Encapsulamento

Como é difícil saber o que realmente será de importância com o passar dos anos, a estratégia de encapsulamento faz com que o objeto digital se mantenha inalterado até que seja necessário. Nessa estratégia mantem-se toda informação necessária a respeito do objeto para que futuramente possa ser desenvolvidos conversores, emuladores ou visualizadores (FERREIRA, 2007).

O encapsulamento nada mais é do que manter o objeto digital exatamente como ele é no presente, para que no futuro apenas sejam desenvolvidos métodos de acessá-los e não de modificá-lo para outra tecnologia.

3.1.4 Refrescamento

A necessidade de manter um objeto digital através do tempo encontra algumas dificuldades e uma delas é referente as delimitações que se encontram nos objetos físicos, o refrescamento faz com que objeto digital sempre possa ser acessado, pois realiza a transferência da informação de um suporte físico a outro mais atual, antes que o outro se torne inacessível por causa do tempo ou pela obsolescência da tecnologia (FERREIRA, 2007).

3.1.5 Conversão para Formatos Concorrentes

Essa estratégia trata-se de converter o objeto digital para um formato que não seja propriamente o da empresa que desenvolveu o *software* proprietário no qual o mesmo foi produzido. Esse tipo de método está restrito apenas a alguns tipos de objetos. O que procurou-se nessa estratégia é guardar informações referentes a descontinuidade do *software*, quando o mesmo não passar por versões atuais (CUNHA, 2007).

3.2 Modelo Conceitual *Open Archival Information System* (OAIS)

O modelo OAIS é um modelo conceitual que tem como função a identificação de componentes funcionais que estarão incluídos em um sistema de informação direcionado a preservação digital. Modelo de referencia desenvolvido através de um processo aberto e iterativo. Esta conceituação pode ser equiparada com duas funções principais de um repositório: a preservação da informação, ou seja, permitir que ela dure um longo período; e proporcionar o acesso à informação de uma forma que atenda as necessidades dos principais usuários do OAIS (LAVOIE, 2004).

Segundo Thomaz (2004) o modelo OAIS é um esquema conceitual que disciplina e guia um sistema para um processo de preservação e manutenção do acesso ao conteúdo digital em um longo prazo. Com o intuito de conscientizar e promover a compreensão de conceitos referentes à preservação de objetos digitais.

Na Figura 6 pode ser visto o modelo de referencia OAIS, com seus diferentes componentes funcionais.

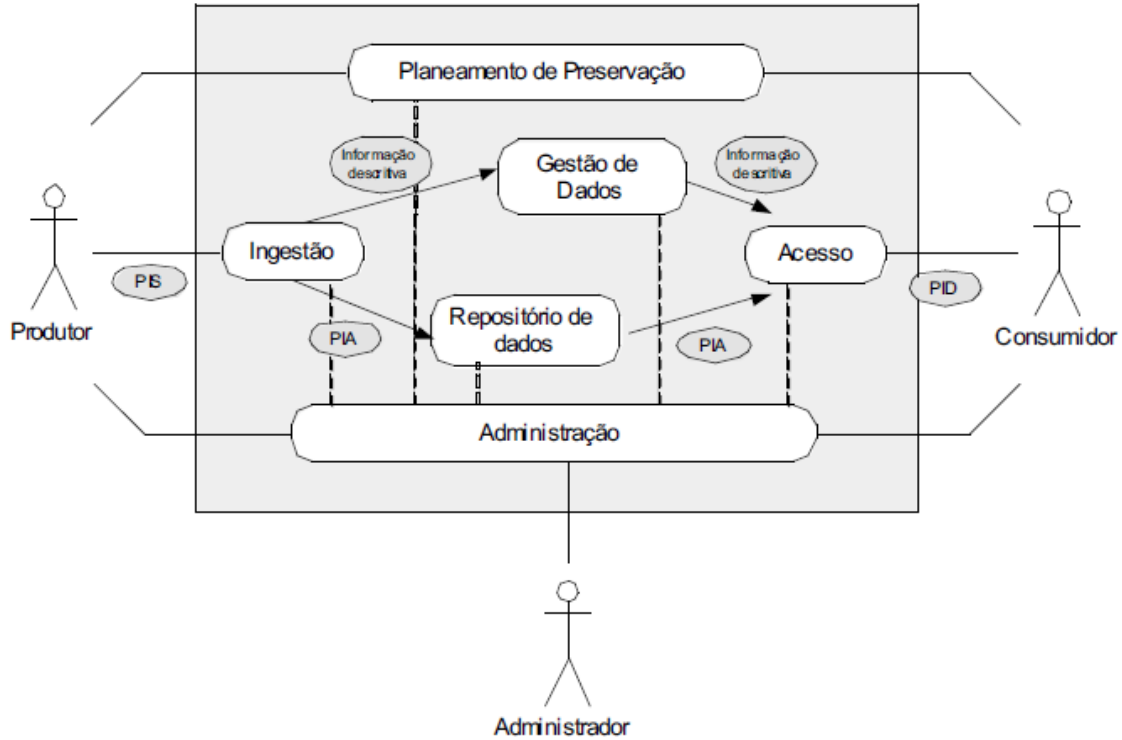


Figura 6 - Modelo de Referência Open Archival Information System (OAIS)
 Fonte: <http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>

O produtor é aquele que produz o material que será preservado, ou seja, o autor da obra que será armazenada em uma biblioteca.

A parte de ingestão é responsável por garantir a integridade da informação que esta sendo recebida.

A gestão de dados é responsável por armazenar a informação descritiva produzida na ingestão.

Repositório é o local em que será armazenado o objeto digital que se deseja preservar.

O componente Planeamento de Preservação é encarregado de definir políticas de preservação, responsável por monitorar o ambiente externo ao repositório e acionar eventos que garantam a preservação quando necessário.

O acesso é o meio entre o repositório e comunidade que tem interesse nos objetos preservados.

A administração é responsável pelas funções diárias de manutenção e também a monitorização dos processos no repositório (FERREIRA, 2006).

3.3 Repositórios Digitais

Segundo a definição do “*Digital Repositories JISC Briefing Paper (2005)*” Repositório Digital é o local onde conteúdos digitais são armazenados, e podem ser pesquisados e recuperados para um uso futuro. Um repositório suporta mecanismos de importação, exportação, identificação, armazenamento e recuperação de recursos digitais. O armazenamento em repositórios digitais permite que as instituições possam melhor gerenciar e preservar seus Objetos Digitais. Os repositórios digitais podem ser utilizados para armazenar artigos de periódicos, teses, e-learning, objetos e materiais de ensino, entre outros (JISC, 2005).

Alguns dos repositórios digitais mais utilizados e confiáveis que atendem aos requisitos tecnológicos são o DSpace, *Flexible Extensible Digital Object and Repository Architecture* (Fedora), *Dark Archive In The Sunshine State* (DAITSS), *Berkeley Electronic Press* (BEPRESS), entre outros (THOMAZ, 2007).

Segundo Café (2003) existem repositórios temáticos e institucionais. Os temáticos reúnem uma coleção de trabalhos de pesquisa que estão direcionados para determinada área do conhecimento. Esse tipo de repositório utiliza tecnologias abertas, fazendo que gerem uma maior acessibilidade a produção de pesquisadores. Os repositórios institucionais agrupam todos os repositórios temáticos de uma organização. No caso de uma instituição de ensino, cada departamento possui seu repositório temático, a união de todos esses repositórios formará o repositório institucional, que pode ser caracterizado como multidisciplinar. Sua principal função é preservar e disponibilizar o conteúdo intelectual que a instituição possui.

Os repositórios tanto temáticos quanto institucionais, possuem determinadas características, com uma estrutura comum de submissão e acesso as informações, são desenvolvidos de acordo com alguns padrões de interoperabilidade específicos. Possuem um grande potencial de agregação e disseminação da informação eletrônica. Algumas das características que os repositórios possuem são determinadas conforme os padrões que utilizam (MONTEIRO, 2007).

Atualmente quando se inicia a construção de uma biblioteca digital uma das questões a ser definida é a utilização de um repositório e de qual repositório.

3.3.1 DSpace

DSpace é um *software* que as organizações acadêmicas tem escolhido bastante para uma melhor manipulação e armazenamento de seus acervos digitais. O código é aberto e não possui fins lucrativos, ou seja, é gratuito e pode ser moldado de acordo com as necessidades da instituição. Permite um ambiente de fácil acesso e é aberto a todos os tipos de conteúdos digitais, que incluem: texto, imagens, imagens em movimento, mpegs e conjuntos de dados. (DSPACE.ORG Acesso em: 10 de junho de 2012)

Algumas órgãos possuem acervos que estão a algum tempo em funcionamento e que foram desenvolvidos utilizando o DSpace, como a Biblioteca Digital da Universidade Federal de Minas Gerais (UFMG), Biblioteca Digital da Câmara dos Deputados, Fundação Getúlio Vargas (FGV), entre outras.

3.3.2 *Flexible Extensible Digital Object Repository Architecture (Fedora)*

Fedora é uma plataforma que possui código aberto desenvolvido pela Universidade de Virginia e Universidade de Cornell, possui uma arquitetura projetada para servir de base para o desenvolvimento de repositórios digitais para uma enorme diversidade de aplicações, como as bibliotecas digitais, repositórios institucionais e outros sistemas gerenciadores de informação (FEDORAPROJECT.ORG Acesso em: 10 de junho de 2012).

A parte central do Fedora é o repositório de serviços, que pode se comunicar através de *Web Services*, permitindo a criação, gerenciamento, armazenamento, acesso e reuso dos objetos digitais. Todas as funções que a plataforma disponibiliza tanto no nível da administração como no nível dos acessos aos objetos digitais estão disponibilizados por este repositório de serviços (OLIVEIRA, 2011).

3.4 Bibliotecas Digitais

Segundo Landoni (1994), a história da biblioteca pode ser dividida em três períodos: a biblioteca tradicional criada por Aristóteles, que vai até a automação de bibliotecas; a biblioteca moderna que tem como função auxiliar o computador para

catalogação e organização do acervo; e a biblioteca eletrônica ou digital que pode ser descrita como a biblioteca do futuro, que surge como uma forma de resgatar as informações, onde o texto completo está à disposição (LANDONI et al. apud MARCHIORI, 1997).

A biblioteca digital na literatura tem diversas definições como biblioteca virtual e biblioteca eletrônica. A biblioteca eletrônica segundo Marchiori (1997) leva a ideia não de uma biblioteca que seja totalmente no computador, mas uma biblioteca que utilize funções do computador para organizar, catalogar, facilitar a busca de texto, entre outras. A biblioteca virtual ela diz como só existindo na realidade virtual, Um *software* encontrado em um computador que proporcione um ambiente que reproduza uma biblioteca, para que ao entrar no ambiente virtual tenha a sensação de estar em uma de verdade.

Segundo Rosetto (1997) ele toma a biblioteca eletrônica da definição anterior como biblioteca automatizada ou organização moderna, que com o crescente aumento da produção de conhecimento e as transformações que aconteceram na sociedade, necessitou-se de processos que se adequassem permitindo o acesso mais rápido a informação bibliográfica. A biblioteca virtual ele toma como sendo em partes como a biblioteca tradicional que existe há vários anos, só que não mais como apenas “depositária”, mas também como um ponto de acesso, tudo isso na rede e com a propriedade de ser acessada de vários lugares.

Na visão de Levacov (1997) existem várias definições para biblioteca virtual, a troca de informações através de mídias eletrônicas, utilizando uma diversidade de aplicativos, dos que utilizam simples caracteres ASCII, a dados que se baseiam em tempo como vídeo, áudio, animações, entre outros. Também uma possibilidade de se concretizar o projeto Xanadu, que é a criação de uma rede que possa ser utilizada com um grande repositório de todos os documentos que a humanidade possui. Esses documentos poderiam apontar para outros que fossem particularmente parecidos ou que possam ser associados, todos digitais e hipertextuais.

A diferença que existe entre a biblioteca virtual e a eletrônica é que a primeira não necessita de interação física para a utilização de seus serviços, enquanto a segunda necessita dessa interação, mesmo utilizando de artifícios da tecnologia para melhor

administração. Biblioteca digital seria uma versão digitalizada da biblioteca original, que pode ser acessada de vários lugares.

A biblioteca digital oferece uma grande variedade de possibilidades na transmissão de ideias, imagens e sons que pode ser acessado por um número ilimitado de pessoas, em qualquer lugar a qualquer tempo (CAUTELA, 2009).

A criação da biblioteca digital foi de grande auxílio, pois com ela pode se manter documentos por um longo período sem ter que se preocupar com aspectos como a deterioração do papel por exemplo. A digitalização aparece junto com vários outros avanços tecnológicos como do cinema para o DVD, ela surge para mudar a cultura atualmente existente da leitura do livro impresso. Como estamos numa sociedade da informação é importante que existam maneiras de preservar os documentos existentes (SODRÉ, 2007).

A sociedade da informação em que vivemos está em constante evolução tecnológica onde as mudanças ocorrem muito rapidamente e com isso as bibliotecas digitais vem com a finalidade de disponibilizar serviços diferenciados das bibliotecas tradicionais. Para atingir esses objetivos elas utilizam cada vez mais ferramentas informacionais que auxiliam na melhor comunicação entre os usuários que estão ao redor do mundo, até mesmo as crianças têm sido atingidas por essas ferramentas informacionais, que surgem mudando a forma como as pessoas buscam as informações que necessitam no seu dia-a-dia (SÁ, 2009).

Segundo Cautela (2009) com o surgimento da *Word Wide Web* e do grande crescimento da internet como uma ferramenta que auxilia no processo de comunicação, nesses últimos anos a recuperação e o acesso às informações se tornou muito mais fácil antes nunca imaginado. As bibliotecas sempre em evolução e transformação constante, buscam acompanhar as mudanças sociais convivendo com o analógico e o digital, o concreto e o virtual, tentando a evolução se não no mesmo ritmo que a tecnologia, bem próxima.

3.5 Serviços de Preservação Digital

Nas seções seguintes serão apresentados alguns serviços que podem ter sua utilização no processo de preservação digital, tais como os serviços providos pelo PRONOM, *Conversion and Recommendation of Digital Object Formats* (CRiB), *JSTOR/Harvard Object Validation Environment* (JHOVE).

3.5.1 Serviço Provido Pelo PRONOM

PRONOM é um registro online de informações técnicas, um recurso para qualquer um que necessite de uma informação imparcial e definitiva sobre os formatos dos arquivos, produtos de *softwares*, e outros componentes técnicos que têm a necessidade de garantir um acesso aos registros eletrônicos e outros objetos digitais que possuem valor cultural, histórico ou de negócios em longo prazo.

A primeira versão do PRONOM foi desenvolvida pelo Arquivo Nacional no departamento de preservação digital em março de 2002. Teve sua criação pela necessidade de um local em que informações técnicas confiáveis a respeito da natureza de registros eletrônicos fossem armazenadas (PRONOM Acesso em: 28 de junho de 2012).

O PRONOM *Persistent Unique Identifier* (PUID) é um esquema utilizado para fornecer persistentes, identificadores únicos e inequívocos para registros no Registro PRONOM. Esses identificadores são de muita importância para o gerenciamento e troca de objetos digitais, fazendo com que os agentes humanos ou automatizados para identifiquem inequivocamente, e que a identificação possa ser utilizada para realizar uma representação necessária para suportar o acesso a um objeto.

O serviço de identificação de objeto que o PRONOM disponibiliza pode ser acessado através do endereço “<http://www.nationalarchives.gov.uk/pronom/Services/Contract/PRONOM.wsdl>”, lugar em que esta situado o documento WSDL do serviço. Acontece uma conexão com o protocolo SOAP através do endereço “<wsdlhttp://www.nationalarchives.gov.uk/pronom/service.asmx?WSDL>” que utiliza a porta “PronomServiceSoap”.

O termo porta citado acima se refere às portas *Transmission Control Protocol (TCP)* e *User Datagram Protocol (UDP)*, que são meios pelos quais *softwares* ou serviços do sistema operacional utilizam para realizar a comunicação em rede, e por consequente, na internet (ALECRIM, 2007).

O serviço provido possui dois métodos o “getSignatureFileVersionV1In” e o “getSignatureFileV1In”, o primeiro retornando uma lista de formatos atualizada sem que seja passado nenhum parâmetro e o segundo passando a versão e uma variável do tipo booleana.

O serviço disponibilizado pelo *Web Service* tem uma lista de formatos, quando acontece a solicitação do serviço um identificador é enviado e assim faz uma busca em sua lista e retorna a quem solicitou o serviço as características referentes ao identificador que foi solicitado.

Cada formato tem um identificador próprio assim como suas versões. Para uma identificação mais eficaz e correta do formato o mais indicado é se utilizar da assinatura interna ao invés da externa (termos que serão explanados nos tópicos subsequentes), pois a assinatura externa verifica a extensão do objeto e ela pode ser facilmente trocada pelo usuário. Com esse tipo de verificação pela extensão pode não se obter exatamente o correto, pois, além disso, a versão em que se encontra o formato não pode ser identificada.

Uma versão simples do modelo de dados que o PRONOM utiliza para estabelecer a relação entre os formatos e suas respectivas assinaturas, pode ser visto no diagrama de classe UML seguinte:

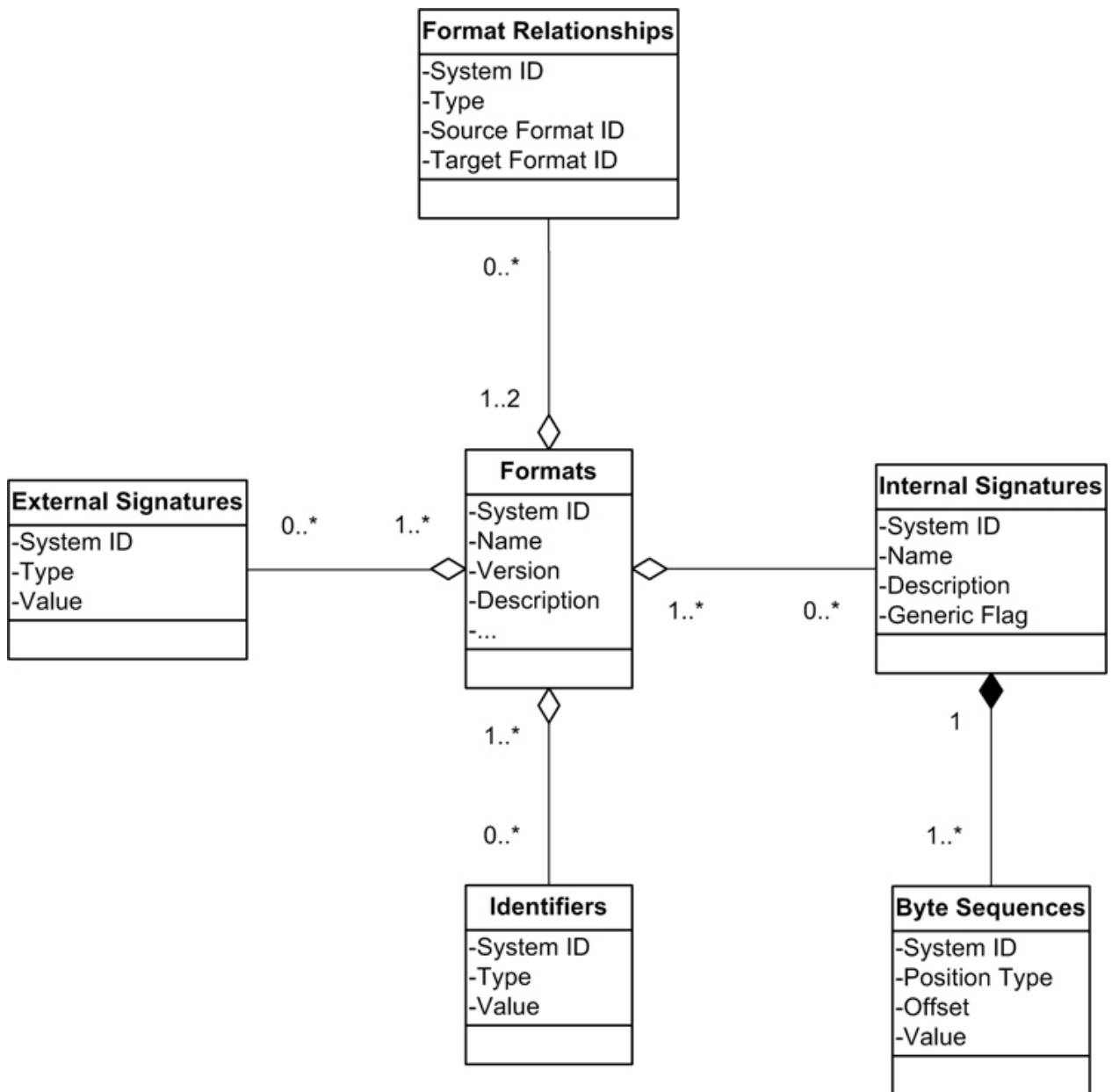


Figura 7 - Esquema de Identificação de Assinatura Externa e Interna
 Fonte: Fonte: (BROWN, 2006)

3.5.1.1 Assinatura Externa

Pode se dizer que a assinatura externa engloba todos aqueles indicadores de formatos que são externos ao *bitstream* do objeto. Em alguns dos sistemas operacionais como o DOS e o *Windows* a assinatura externa é fornecida pela extensão que possui o arquivo, por exemplo, as extensões doc do *Microsoft Office Word*. A real

utilização da extensão não serve propriamente para realizar a identificação de formato, mas sim dizer ao sistema operacional qual pacote de *software* deve ser utilizado para abrir o objeto. A assinatura externa serve mais para que se possa ter uma visão geral sobre o formato do arquivo.

3.5.1.2 Assinatura Interna

A assinatura interna engloba todos os indicadores de formato que estão incluídos no *bitstream* do objeto. Cada especificação de formato tem uma forma específica de organizar o conteúdo, isso faz com que todos os objetos desse formato possuam essa mesma característica, fazendo que esse tipo de identificação seja mais eficaz que a assinatura externa. Nos serviços providos pelo PRONOM a assinatura interna é composta por uma ou mais sequência de *bytes*, todas elas possuindo uma sequência consecutiva de valores hexadecimais. Os *bytes* necessários para a identificação do objeto podem ser colocados de duas formas: absoluta, que a sequência de *bytes* começa numa posição fixa dentro da *bitstream* e variável, que a sequência de *bytes* pode começar em qualquer deslocamento dentro da *bitstream*.

3.5.2 Conversion and Recommendation of Digital Object Formats (CRiB)

O CRiB é uma arquitetura orientada a serviços (SOA), projetado para auxiliar instituições que possuem patrimônio cultural na migração dos documentos do formato físico para o digital, preocupando-se principalmente com a preservação desses documentos. O sistema do CRiB funciona avaliando as conversões realizadas nos documentos, tendo como padrão de qualidade as exigências de cada instituição cliente (CRiB Acesso em: 29 de junho de 2012).

O serviço web CRiB disponibiliza um conjunto de métodos que permite que a aplicação cliente possa realizar atividades como, migrações de formatos complexos, determinar a quantidade de dados perdidos na migração, as intervenções de preservação de documentos e a obtenção de alternativas de migrações para que se possa preservar adequadamente os acervos de objetos digitais (FERREIRA, 2006).

Na Figura 8 pode ser vista a arquitetura do CRiB, na camada de aplicação representa os clientes que utilizarão o serviço e os repositórios que armazenarão os objetos. A camada de negócios representa o sistema do serviço, com seu sistema de

registro que identifica o serviço de conversão, o corretor de migração lida com migrações locais e remotas, o avaliador de objeto que detecta perdas durante a migração, o avaliador de formatos que disponibiliza informações sobre o estado atual do objeto e o conselheiro de migrações que junta todas as informações do objeto para gerar sugestões de migração.

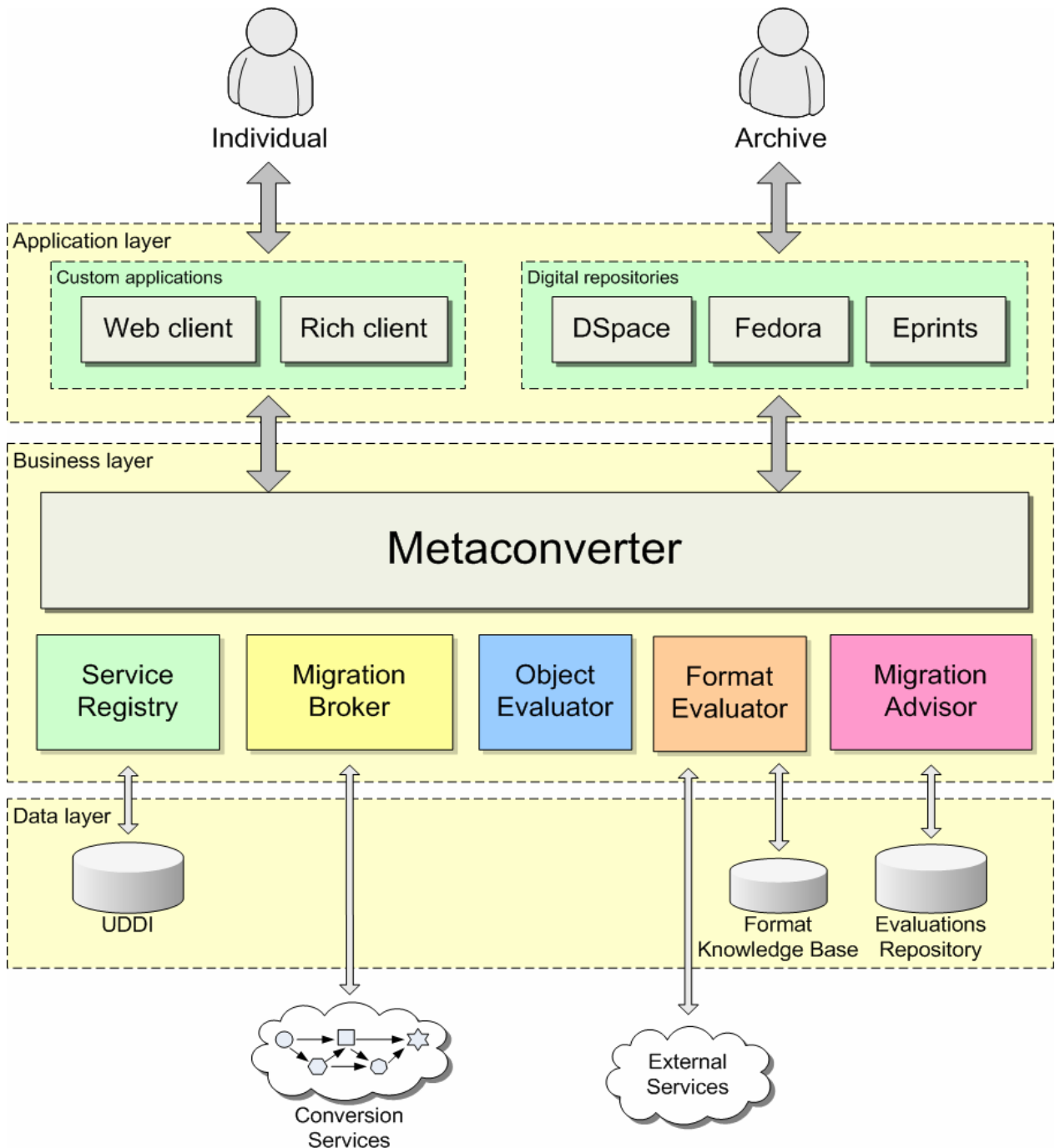


Figura 8 - Arquitetura CRiB

Fonte: <http://repositorium.sdum.uminho.pt/bitstream/1822/4457/1/XATA06-0.11.pdf>

3.5.3 JSTOR/Harvard Object Validation Environment (JHOVE)

O JHOVE é um serviço de validação de formato desenvolvido para possuir sua estrutura extensível, para que repositórios possam se tornar capazes de automatizar seus procedimentos na medida do possível. O serviço possui funções como a identificação de formato, que é o processo de verificar se o formato do objeto digital esta em conformidade, a validação do formato, para analisar se o suposto formato que o objeto digital possui esta em conformidade com o real formato, e a caracterização do objeto, que é a extração das propriedades do formato especifica do objeto (HUL.HARVARD.EDU Acesso em: 29 de junho de 2012).

Na Figura 9 pode ser identificada a utilização do JHOVE, primeiramente na seção *Deposit agent* ele verifica o objeto digital e extrai os metadados, os mesmos junto com o objeto são dirigidos ao repositório. Na seção *Object repository* o objeto digital é separado dos seus metadados e uma nova extração dos metadados do objeto acontece, para confirmar se os metadados que vieram com o objeto estão de acordo com os recém-extraídos e assim realizar a ingestão no repositório.

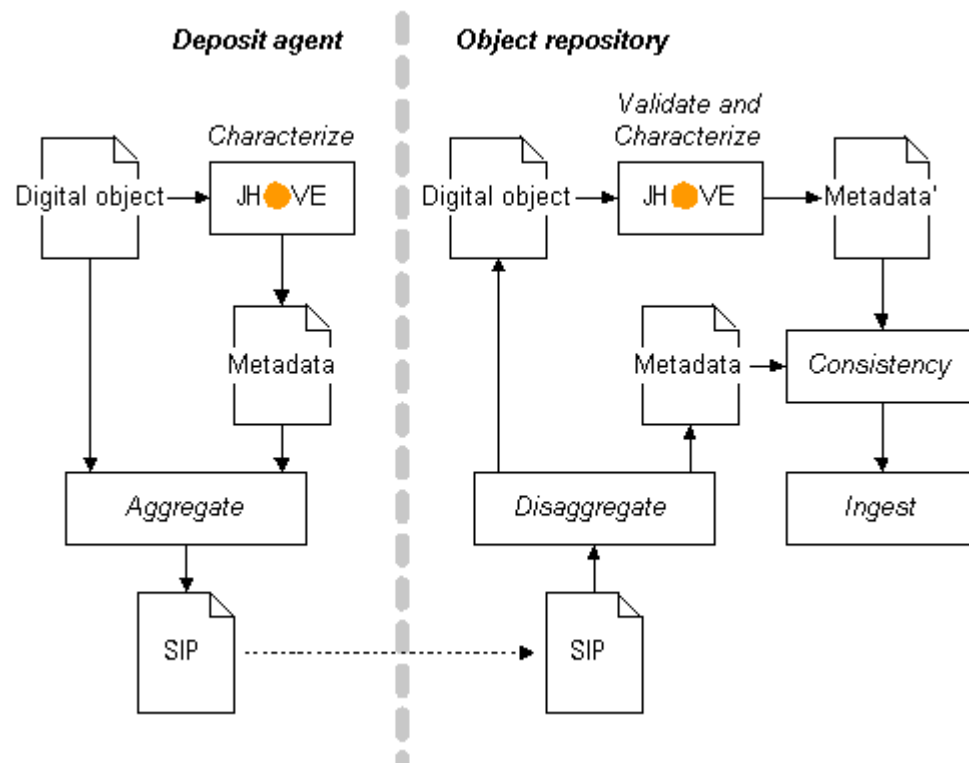


Figura 9 - Utilização do JHOVE
Fonte: <http://hul.harvard.edu/jhove/>

4 PROJETO

Este projeto propõe o desenvolvimento de uma aplicação que utilize serviços web disponíveis na internet e que possam ser utilizados no processo de preservação digital. Antes que o objeto digital possa ser dirigido aos respectivos repositórios, eles precisam passar por alguns métodos de identificação do objeto ou de modificações para um formato padrão. A aplicação será necessária antes de o objeto ser dirigido ao repositório, para que futuramente o arquivo possa ser utilizado sem que perca alguma de suas características.

O serviço utilizado nesse projeto foi o provido pelo PRONOM, mas outros serviços foram identificados e não puderam ser utilizados. O *Conversion and Recommendation of Digital Object Formats* (CRiB), foi um dos primeiros a ser provavelmente utilizado, mas no momento do desenvolvimento o serviço não estava disponível. O serviço de identificação de formatos JHOVE desenvolvido pela *Harvard University*, também foi um dos pesquisados, mas não foi utilizado pela indisponibilidade do documento WSDL do serviço.

4.1 Material

4.1.1 Linguagem de programação

A linguagem de programação utilizada nesse projeto foi Java que é orientada a objeto, desenvolvida pela *Sun Microsystems*. Foi escolhida por ser multiplataforma, abrangendo uma gama maior de dispositivos. A versão do Java utilizada foi a 1.6.0_32 que esta presente no pacote Java SE 6 Update 32.

4.1.2 API's e Bibliotecas

A API utilizada dá suporte a serviços web foi a Java API for XML *Web Services* (JAX-WS), que é de fácil utilização para aplicações que utilizam XML. Disponibilizando o documento WSDL do serviço web utilizado com essa API, o trabalho com os métodos desse serviço ficam muito mais fáceis.

Outra API foi a Apache Tika, que é utilizada para extrair metadados de arquivos, ela auxiliou na facilitação da leitura do objeto digital, que após isso teve seu *mimetype* identificado.

4.1.3 Ambiente de Desenvolvimento

O ambiente de desenvolvimento utilizado foi o NetBeans por ser principalmente direcionado a linguagem de desenvolvimento Java e também por ser de maior facilidade de programação do autor desse trabalho.

4.2 APWS - Aplicação de Serviços Web

O que se propõe nesse trabalho é a criação de uma aplicação para consumo de serviços web. A aplicação desenvolvida em Java tem uma estrutura de fácil entendimento e utilização. Ela está posicionada para utilização antes do objeto digital ser dirigido aos repositórios, para que se tenha um melhor conhecimento sobre ele e assim possa ser armazenado de forma eficiente sem perder suas características, essas que poderão ser utilizadas posteriormente para a migração dele para outra tecnologia.

A aplicação utiliza os serviços providos pelo PRONOM, que faz a identificação sobre o real formato do objeto, verificando tanto sua assinatura externa quanto a sua assinatura interna, fazendo assim uma seleção das propriedades do objeto como, nome do arquivo, código identificador, código único do PRONOM, *mimetype*, entre outros. A figura a seguir mostra a tela principal da aplicação:

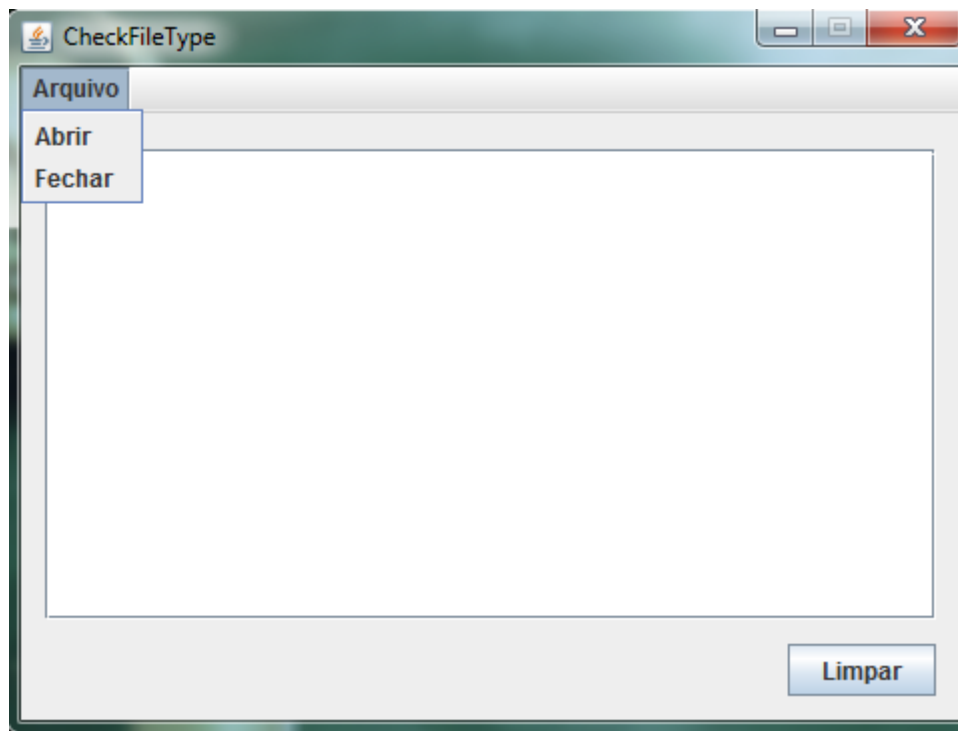


Figura 10 - Tela de Verificação
Fonte: Elaborado pelo Autor

Desenvolvida em Java e de fácil entendimento, ela conta com um menu que possui dois campos “Abrir” e “Fechar”, o campo abrir, abre outra janela que pode ser vista na Figura 11 que servirá para a busca do arquivo e assim encontrar seu diretório.

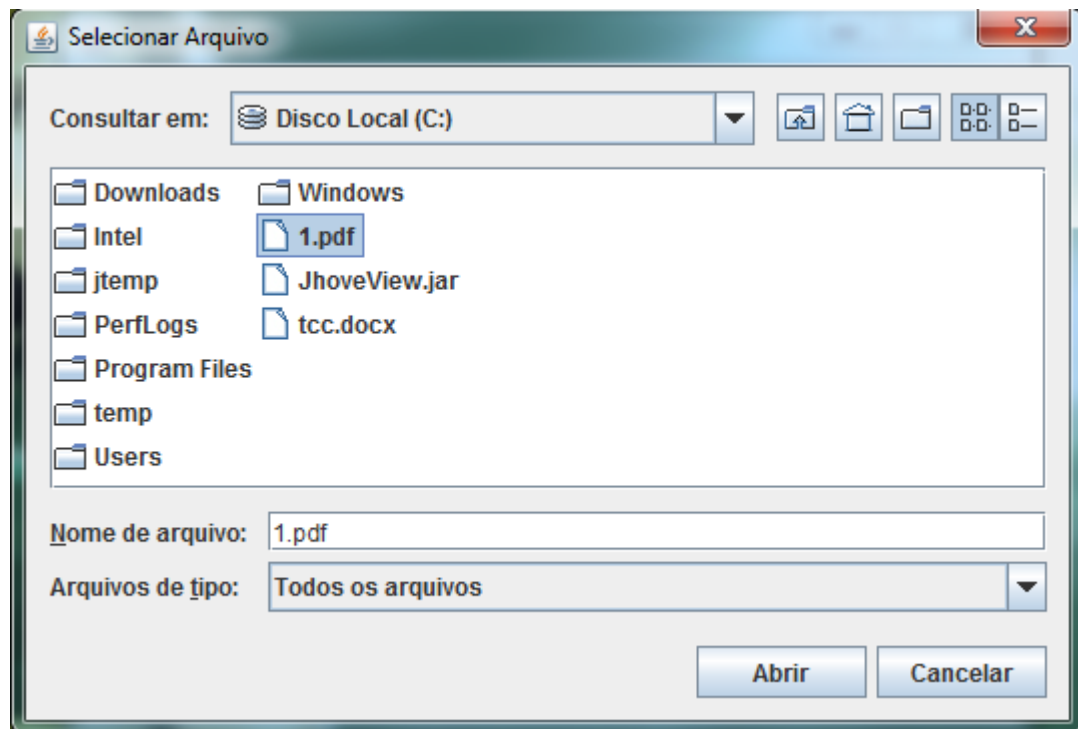


Figura 11 - Tela de Seleção do Arquivo
Fonte: Elaborado pelo Autor

Se nenhum arquivo for escolhido aparece na tela uma mensagem “Arquivo Não Selecionado” como pode ser visto na Figura 12.

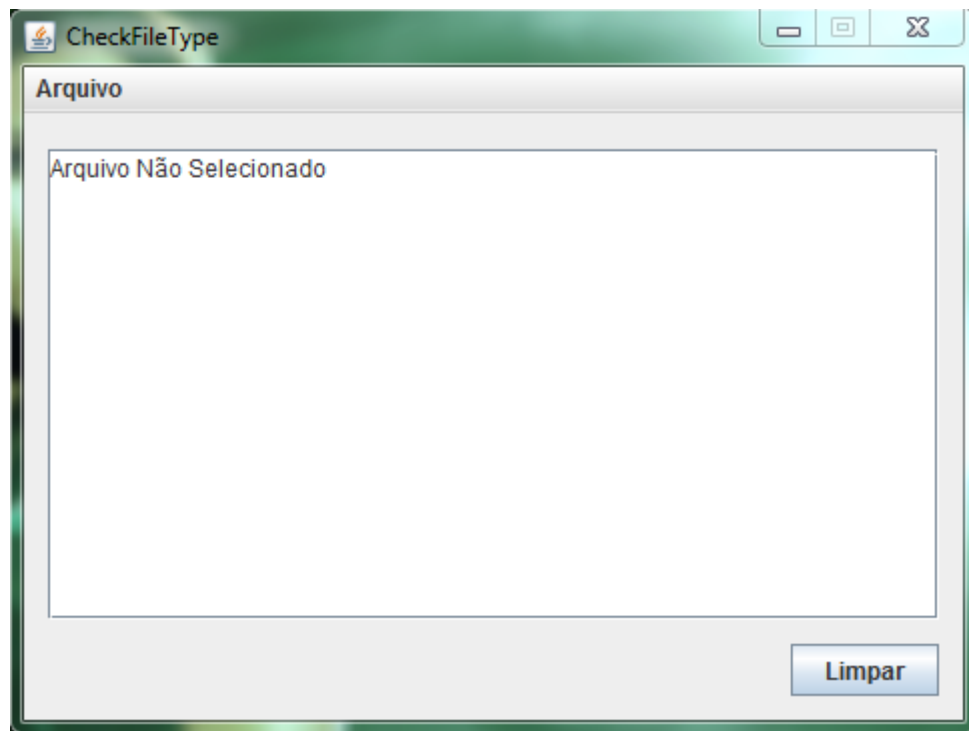


Figura 12 - Arquivo não selecionado
Fonte: Elaborado pelo Autor

Assim que o arquivo é escolhido começa uma verificação do seu tipo, utilizando a lista de formatos provida pelo PRONOM, fazendo uma comparação dos dados do arquivo selecionado com a lista do serviço web, enquanto a aplicação realiza essa operação aparece uma mensagem de “Aguarde...” que pode ser vista na Figura 13, até que a verificação tenha sido realizada.

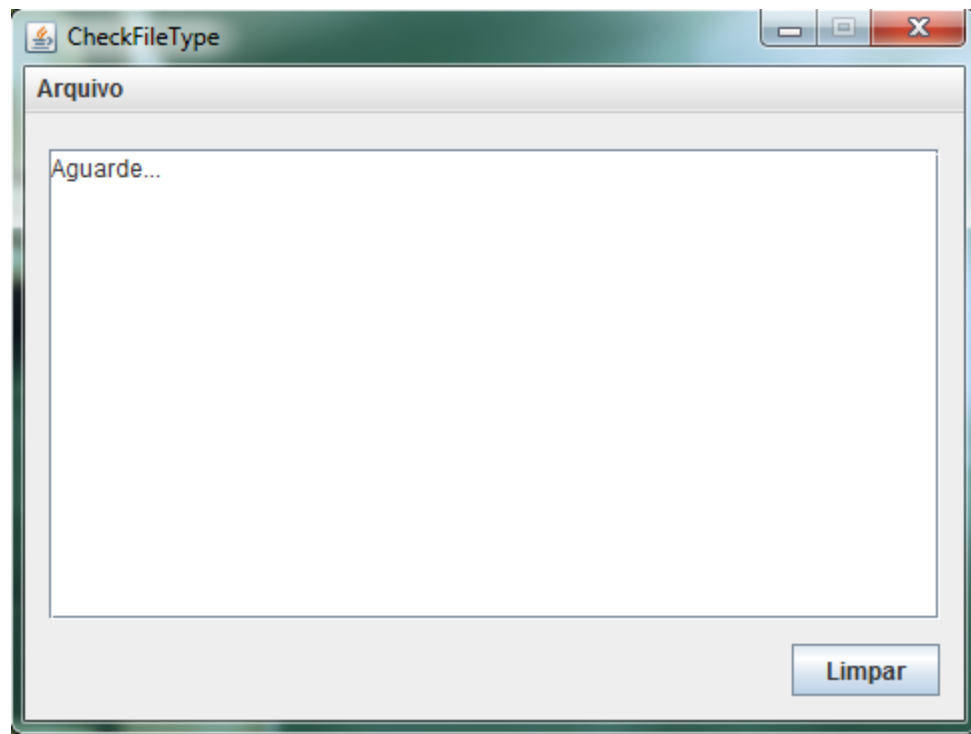


Figura 13 - Tela de Espera
Fonte: Elaborado pelo Autor

Após a verificação é mostrado no campo de texto as características que foram encontradas a respeito do arquivo, como:

File Name: nome do arquivo da mesma maneira em que ele se encontra dentro do diretório.

MIME Type: tipo que o arquivo possui, faz a identificação do arquivo, se é um documento, imagem, apresentação de slides, arquivo de audio entre outros.

Name: nome completo do tipo do arquivo, como no caso do mp3, por exemplo, *MPEG 1/2 Audio Layer 3 Streaming*.

PUID: é um identificar único que o PRONOM utiliza para registrar os formatos de arquivos, diferentemente do esquema do *mimetype* o puid oferece uma identificação mais abrangente.

ID: um identificador para cada formato, no caso da extensão do Microsoft Word todos os arquivos que possuírem a extensão “doc” terão um identificador único para esse formato, já com o docx mesmo sendo da mesma linha será outro por ser como se fosse um novo formato.

Internal Signature: assinatura interna do objeto digital, o real formato que o arquivo possui, sem se preocupar com sua extensão. Como pode ser visto na Figura 14.

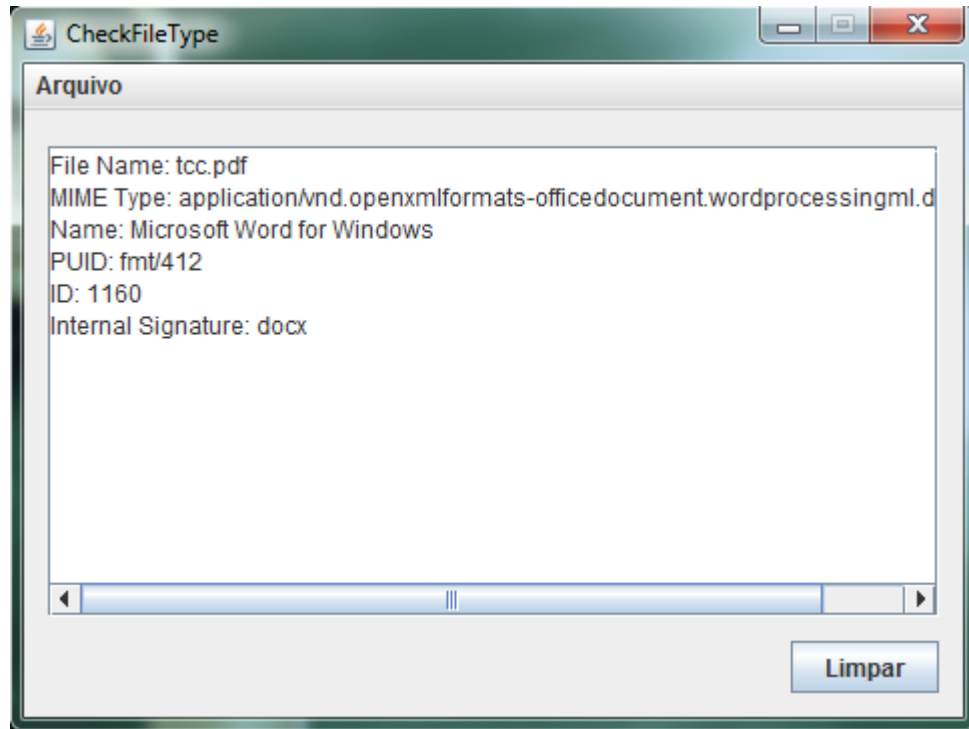


Figura 14 - Tela de Verificação Concluída
Fonte: Elaborado pelo Autor

4.2.1 Modelagem da Arquitetura do Sistema

Nos próximos tópicos são exibidos alguns diagramas UML (Linguagem Unificada de Modelagem) que serão utilizados para a melhor compreensão do sistema proposto.

4.2.1.1 Diagrama de Caso de Uso

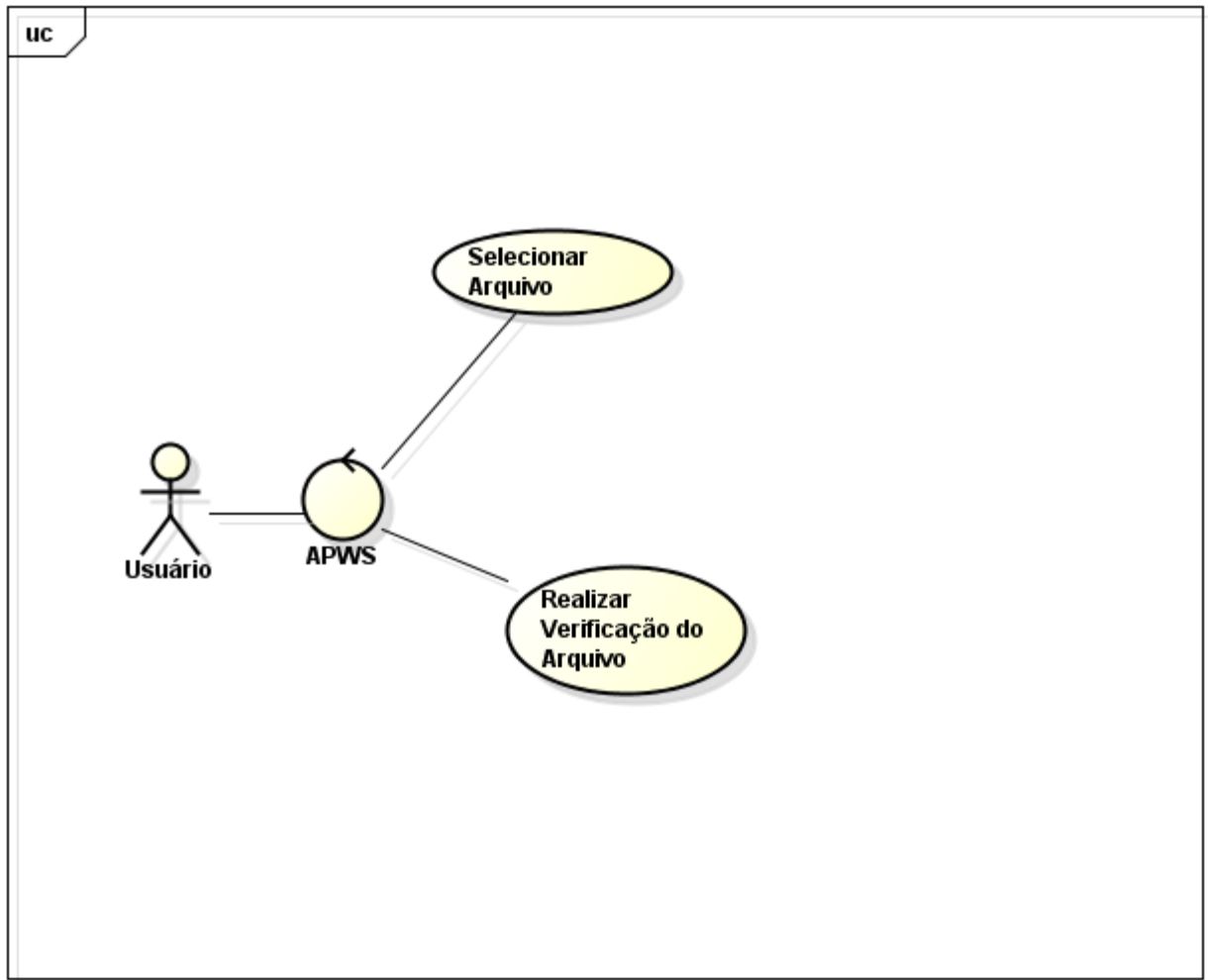


Figura 15 - Diagrama de Caso de Uso
Fonte: Elaborado pelo Autor

Casos de uso = Selecionar Arquivo

Ator = Usuário

Pré-condição = O sistema deve ter sido iniciado.

Pós-condição = O arquivo deve ter sido selecionado.

Fluxo principal =

- 1- O usuário através do menu da aplicação escolhe a opção abrir.
- 2- O usuário busca o diretório em que está o arquivo desejado.
- 3- O usuário utiliza o botão abrir para escolher o arquivo e iniciar a verificação.

Fluxo Alternativo =

2.1 – O usuário cancela a busca.

2.1.1-Uma mensagem aparece na tela de que o arquivo não foi selecionado.

Casos de uso = Realizar Verificação do Arquivo

Ator = APWS

Pré-condição = O sistema deve ter sido iniciado.

Pós-condição = A classe MimeArquivo deve ter tido seus atributos preenchidos.

Fluxo principal =

- 1- O sistema realiza uma análise do arquivo escolhido.
- 2- O sistema armazena o *mimetype* do arquivo.
- 3- O sistema busca no servidor web uma lista que contém uma coleção de formatos.
- 4- O sistema compara o *mimetype* do arquivo escolhido com a lista que recebeu do servidor.
- 5- O sistema com base no formato que foi identificado na lista armazena as informações referentes na classe MimeArquivo.
- 6- O sistema exibi na tela as informações referentes ao arquivo que foi analisado.

Fluxo Alternativo =

2.1 – Se o arquivo possuir algum tipo de não conformidade.

2.1.1-Uma mensagem de erro aparece.

4.1 - Se o *mimetype* do arquivo não estiver na lista.

4.1.1- Uma mensagem de erro aparece.

4.2.1.2 Diagrama de Classe Análise

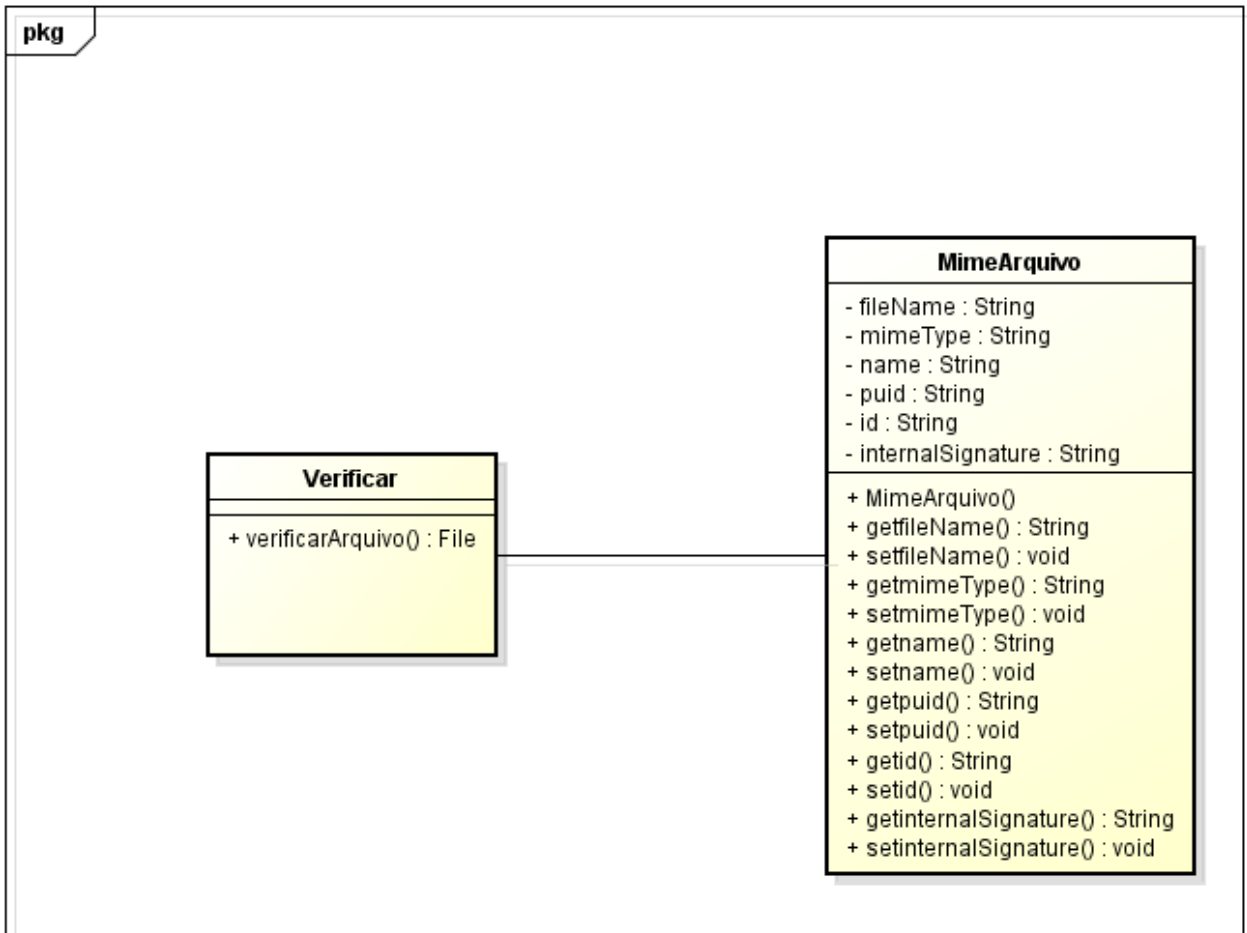


Figura 16 - Diagrama Classe Análise
 Fonte: Elaborado pelo Autor

No diagrama de classe análise podem ser identificadas duas classe, a Verificar que possui apenas o método verificarArquivo, responsável por analisar o objeto digital e através dos seus metadados armazenar o *mimetype* do arquivo selecionado. Depois busca a lista que contém a coleção de formatos no serviço web e compara o *mimetype* do arquivo com todos os existentes na lista. Identificando aquele que é igual, armazena na classe MimeArquivo que é a segunda do diagrama as características que são interessantes para a aplicação como o *fileName*, *mimeType*, *name*, *puid*, *id*, *internalSignature*.

4.2.1.3 Diagrama de Comunicação

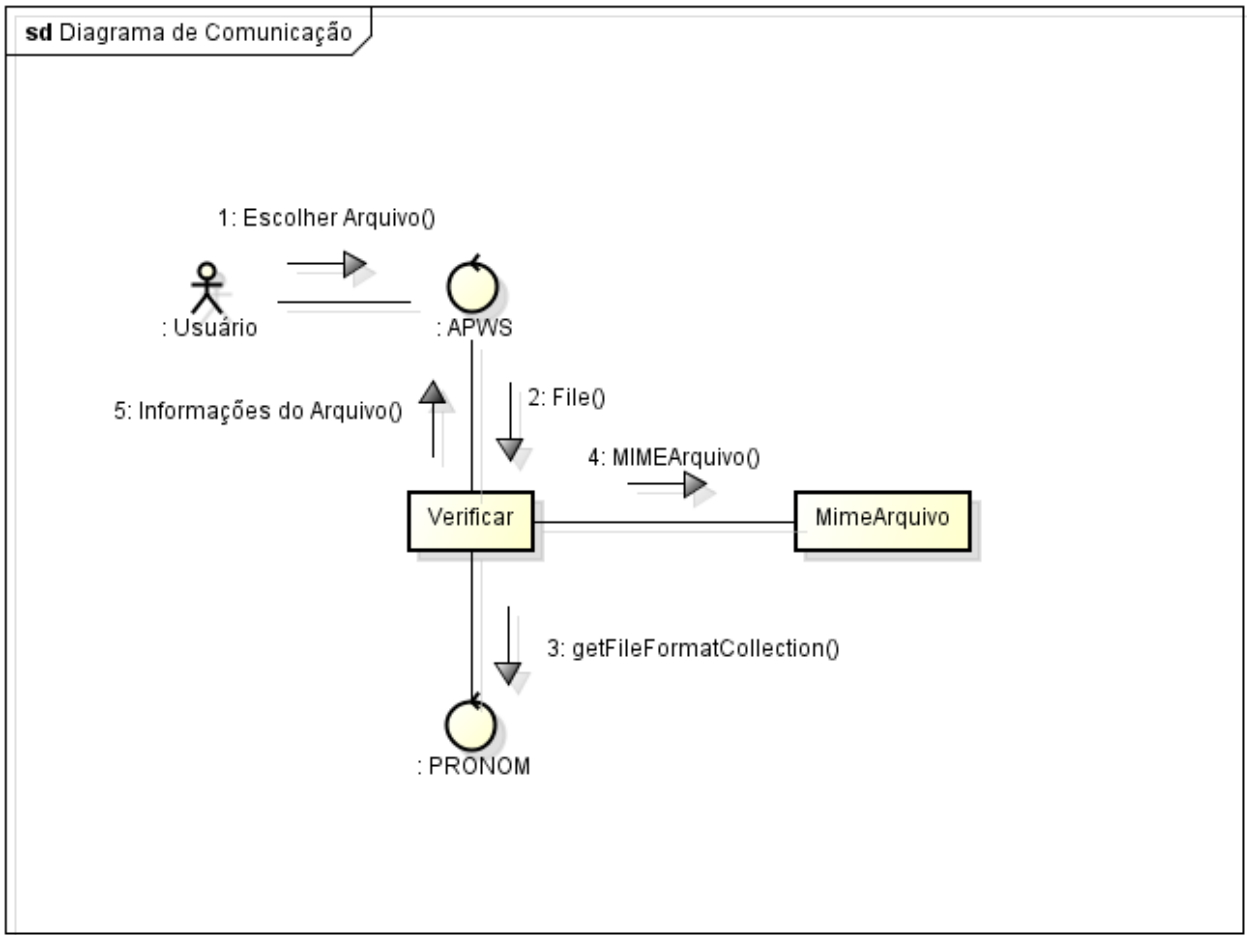


Figura 17 - Diagrama de Comunicação
 Fonte: Elaborado pelo Autor

No diagrama de comunicação pode ser visto grande parte do processo de execução da aplicação, primeiramente o usuário através da interface selecionará o arquivo que será verificado. Na classe Verificar, o arquivo será analisado e com base nos seus metadados armazenado o *mimetype*, será requisitado do serviço web uma lista que contém uma coleção de formatos para que possa acontecer uma comparação do *mimetype* do arquivo com cada um da coleção provida pelo WS. Quando houver a identificação daquele que for igual, a classe armazena na MimeArquivo as propriedades do formato interessantes para a aplicação e retorna ao usuário as informações referente ao arquivo.

4.2.1.4 Diagrama de Sequência

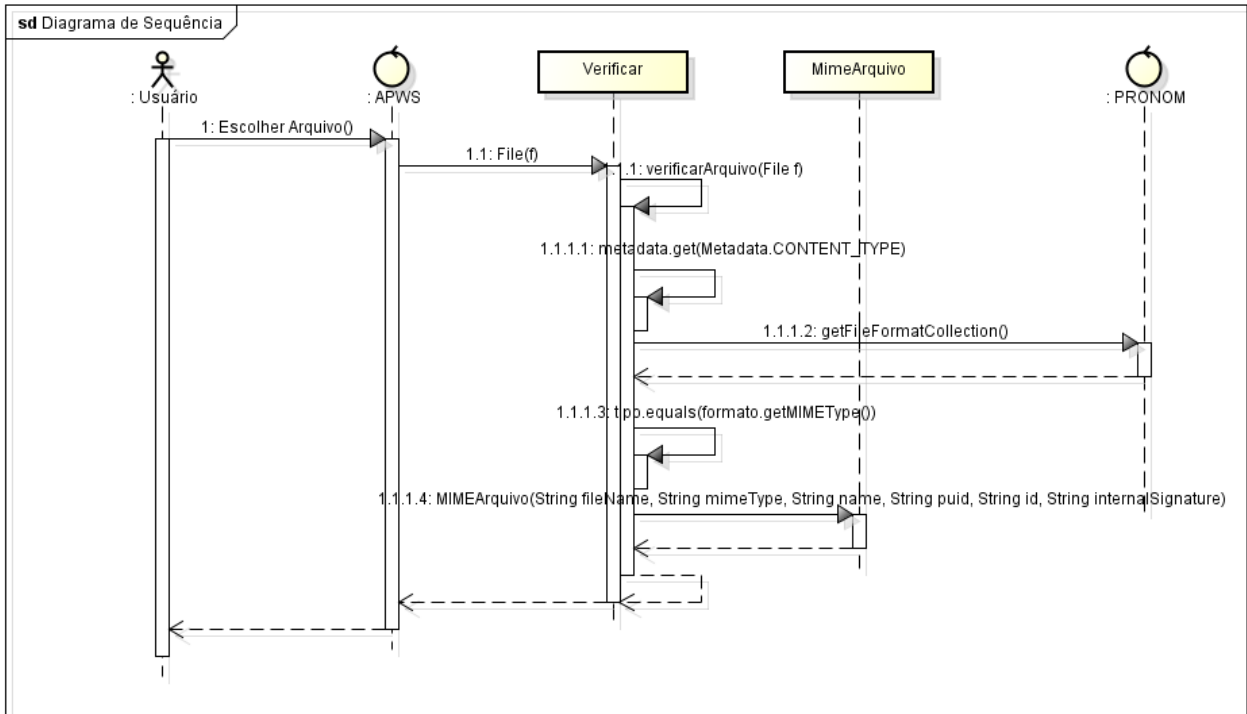


Figura 18 - Diagrama de Sequência
Fonte: Elaborado pelo Autor

O diagrama de sequência não é muito diferente do diagrama de comunicação, o usuário começa escolhendo o arquivo que deseja verificar, a classe Verificar analisa o arquivo e armazena o *mimetype*, depois solicita ao serviço web uma lista que contém uma coleção de formatos, cada item dessa lista é comparado com o *mimetype* do arquivo até que se identifique o igual ao que foi passado. As informações encontradas através da lista são armazenadas na classe MimeArquivo e exibidas na interface para o usuário.

4.2.1.5 Diagrama de Estados Navegacionais

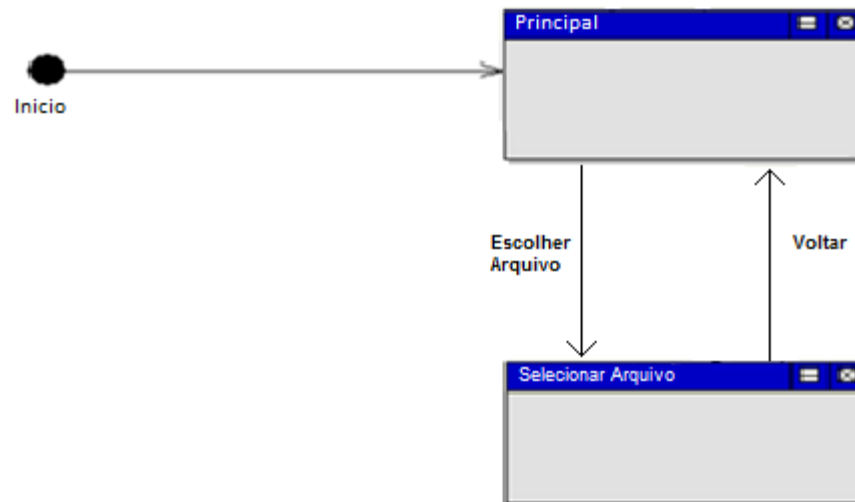


Figura 19 - Diagrama de Estados Navegacionais
Fonte: Elaborado pelo Autor

A aplicação possui apenas duas telas: uma principal e uma de seleção de arquivo. A maioria das operações acontece na tela principal, é nela que através de um menu consegue-se abrir a tela de seleção que é utilizada para buscar o arquivo que deseja realizar a verificação, enquanto o processo é realizado na principal aparece a mensagem "Aguarde" e logo após as informações referentes ao arquivo escolhido.

4.2.1.6 Diagrama de Atividade

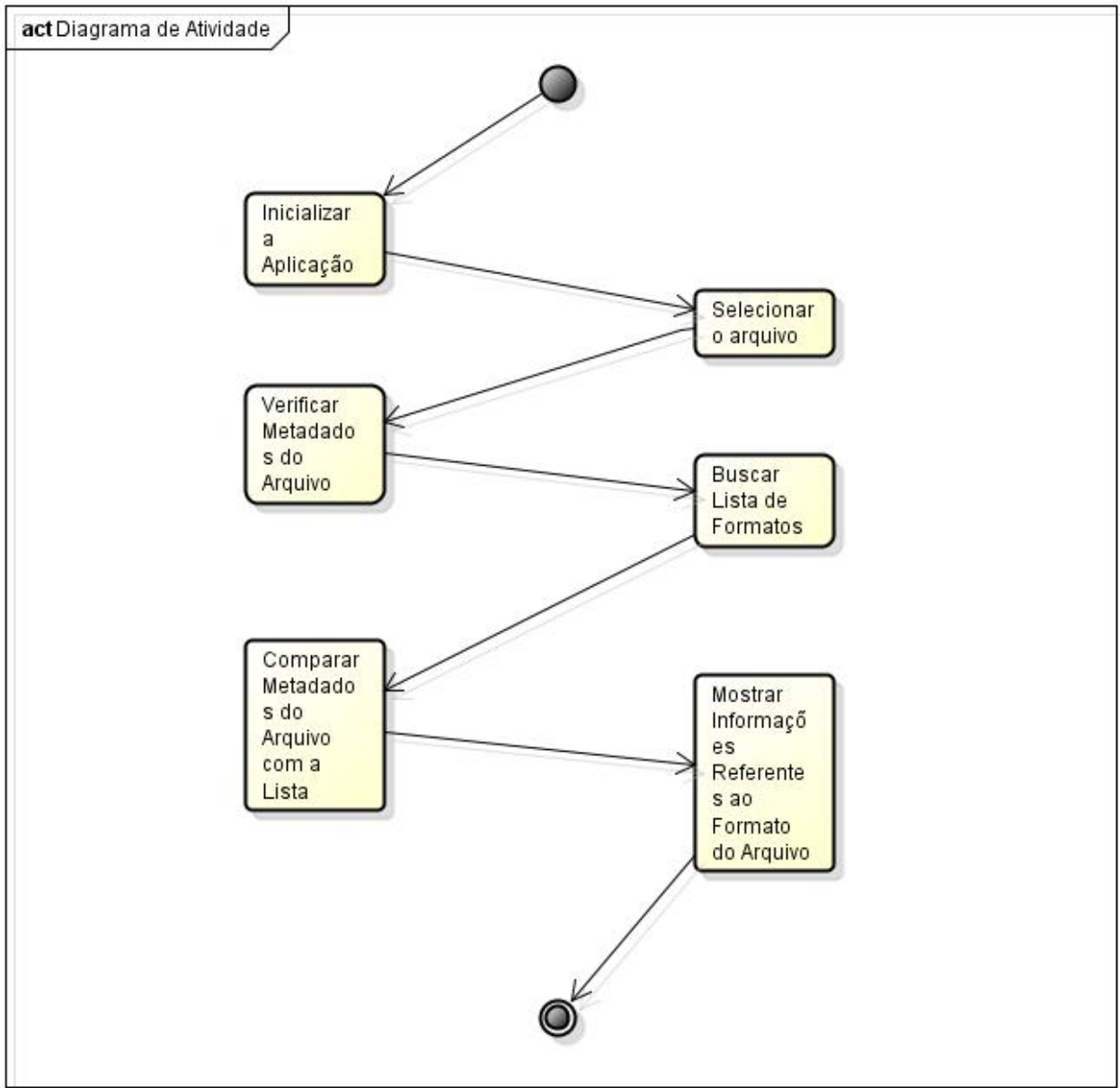


Figura 20 - Diagrama de Atividade
Fonte: Elaborado pelo Autor

No Diagrama de Atividade na Figura 20, pode ser visto a sequência de passos que a aplicação realiza em sua execução. Primeiramente o usuário inicia a aplicação e escolhe o arquivo que deseja ser realizada a verificação, com o arquivo escolhido a aplicação obtém os metadados do arquivo e busca uma lista de formatos no serviço web para realizar a comparação dos metadados do arquivo com cada item da lista.

Identificando aquele que se iguala a algum da lista a aplicação armazena os dados referentes aquele formato na classe MimeArquivo e exibe para o usuário.

5 DISCUSSÃO DE RESULTADOS

O projeto foi desenvolvido tendo em mente auxiliar no processo de preservação digital, realizando a identificação de formatos e colhendo algumas informações referentes ao objeto digital. A aplicação funciona de modo que ao obter as informações, elas possam ser utilizadas futuramente para que não haja complicações na leitura do arquivo.

Na parte de desenvolvimento da aplicação, houve algumas dificuldades com relação ao acesso dos serviços web, muito deles não estavam mais em funcionamento ou o acesso a eles se mostrava muito complicado.

A aplicação não possui uma agregação direta com outras aplicações, mas pode ser utilizada em conjunto, fazendo com que assim o sistema de preservação digital possa se tornar mais eficiente. Ela realiza a identificação de algumas informações para que no futuro essas mesmas informações sirvam para facilitar o processo de emulação ou migração do objeto digital.

6 CONCLUSÃO

O que se procurou buscar nesse trabalho foi o desenvolvimento de uma aplicação que consuma algum serviço web, utilizado no processo de preservação digital, para facilitar a identificação do documento e suas características quando houver o armazenamento em repositórios. O tema foi escolhido pela necessidade de identificar realmente o formato do objeto digital, para que quando ele for utilizado posteriormente, não existir empecilhos em sua execução.

O principal objetivo desse projeto é a seleção de algum serviço que auxiliaria na preservação de documentos, pois com o rápido avanço da tecnologia muitos documentos podem ficar obsoletos por se tornarem de difícil acesso, ou as tecnologias de acesso a eles não serem mais utilizadas. O serviço que foi utilizado foi provido pelo PRONOM, que realiza a identificação do formato do objeto digital.

Foi realizada uma pesquisa bibliográfica sobre biblioteca digital, preservação digital e serviços web e com isso pode se ter uma visão mais ampla sobre o assunto, também foi selecionado um serviço web de preservação digital e entendido seu funcionamento, e por último o desenvolvimento da aplicação que consome o serviço.

A maior dificuldade encontrada nessa pesquisa foi à descoberta dos WSDL's dos serviços web, que são os responsáveis pela conexão da aplicação com o serviço proposto. Alguns serviços foram analisados e não puderam ser utilizados por não ter sido encontrado o documento WSDL e também pelo serviço não estar mais em funcionamento, ou seja, online. O serviço utilizado funcionou de maneira eficaz para o que foi proposto, passando as informações necessárias ele retorna corretamente o formato do objeto digital.

Concluindo, a aplicação funciona de forma eficiente realizando a identificação dos formatos dos objetos digitais, se houvesse mais tempo para a implementação poderia ser feito uma aplicação completa, com todos os serviços necessários, fazendo assim que somente ela fosse utilizada antes dos objetos digitais serem dirigidos aos respectivos repositórios.

REFERÊNCIAS

ALECRIM, Emerson. **Portas TCP e UDP**. Disponível em: <<http://www.infowester.com/portastcpudp.php>>. Acesso em: 28 de junho de 2012.

ARELLANO, Miguel Angel. **Preservação de Documentos Digitais**. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>>. Acesso em: 28 de junho de 2012.

BASIURA, Russ. **Profissional ASP.NET Web Services**. 2003 Editora Pearson Education do Brasil

BROWN, Adrian. **Automatic Format Identification Using PRONOM and DROID, 2006**. Disponível em: <http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/automatic_format_identification.pdf>. Acesso em: 15 de maio de 2012.

CAFÉ, Lígia; et al. **Repositórios institucionais: nova estratégia para publicação científica na Rede**. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26., 2003, Belo Horizonte. Anais... Belo Horizonte: INTERCOM, 2003.

CAUTELA, Lucinda de Jesus Teixeira Campos. **Curso de Especialização em Instituições e Processos Políticos do Legislativo, Câmara dos Deputados, Centro de Formação, Treinamento e Aperfeiçoamento (Cefor)**, 2009.

CHINNICI, Roberto. **Web Services Description Language (WSDL) Version 1.2**. Disponível em: <<http://www.w3.org/TR/2002/WD-wsdl12-20020709/>>. Acesso em: 27 de junho de 2012.

CRIB, **CRiB (Conversion and Recommendation of Digital Object Formats)**. Disponível em: <<http://www.dcc.ac.uk/resources/external/crib-conversion-and-recommendation-digital-object-formats>>. Acesso em: 29 de junho de 2012.

CRUZ, Sérgio Manuel Serra da. **Serviços Web – Uma breve introdução (Parte I)**. Informativo Técnico do Núcleo de Computação Eletrônica da UFRJ - Primeira Edição - Jul/Ago 2005

CUNHA, Jacqueline de Araújo., LIMA, Marcos Galindo. **Preservação Digital: o estado da arte**. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--043.pdf>> Acesso em: 28 de junho de 2012.

DANTAS, Daniel Chaves Toscano. **SOAP (Simple Object Access Protocol)**. Disponível em: <http://www.gta.ufrj.br/grad/07_2/daniel/index.html>. Acesso em: 01 de março de 2012.

DSPACE. **DSpace** . Disponível em: <<http://www.dspace.org/introducing>>. Acesso em: 10 de junho de 2012.

EVDEMON, John. **Princípios do design de serviço: padrões e antipadrões de serviço**. Disponível em: < <http://msdn.microsoft.com/pt-br/library/ms954638>>. Acesso em: 27 de junho de 2012.

FEDORA. **Fedora**. Disponível em: <<http://fedoraproject.org/pt/>>. Acesso em: 10 de junho de 2012.

FERREIRA, José Miguel Araújo. **Solutions Walkthrough Report**. Disponível em: <http://www.digitalpreservationeurope.eu/publications/challenge_reports/ferreira_report.pdf>. Acesso em: 10 de Junho de 12.

FERREIRA, Miguel, BAPTISTA, Ana Alice, RAMALHO, José Carlos, ed. lit. - "**CRIB : a service oriented architecture for digital preservation outsourcing**". Paper presented at the XATA - XML: Aplicações e Tecnologias Associadas, Portalegre, 2006. ISBN 972-99166-2-4.

FERREIRA, Miguel. **Introdução a Preservação Digital: Conceitos, estratégias e actuais consensos**. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 02 de junho de 2012.

JESUS, Joana D'arc Pereira de. **Preservação da informação digital: estudo de caso na Biblioteca Digital de Monografias da Universidade de Brasília**. 2011. xiv, 69 f.

Monografia (Bacharelado em Biblioteconomia)—Universidade de Brasília, Brasília, 2011.

JHOVE, **JHOVE - JSTOR/Harvard Object Validation Environment**. Disponível em: <<http://hul.harvard.edu/jhove/>>. Acesso em: 29 de junho de 2012.

JISC. **Briefing Paper – Higher Education Sector**. Disponível em: <http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository%28HE%29-v1-final.pdf>. Acesso em: 27 de junho de 2012.

KEEFER, Alice. **Preservación digital y depósitos institucionales**. *El Profesional de la información*, 2005b, v. 14, nº 6, nov.-dic., p. 404-406. Disponível em: <<http://www.elprofesionaldelainformacion.com/contenidos/2005/noviembre/1.pdf>>. Acesso em: 27 de junho de 2012.

LAVOIE, Brian F. **The Open Archival Information System Reference Model: Introductory Guide**. Disponível em: <http://www.dpconline.org/docs/lavoie_OAIS.pdf>. Acesso em: 29 de junho de 2012.

LEE, K.-H.; SLATERRY, O.; LU, R.; TANG, X.; McCRARY, V.. **"The State of the Art and Practice in Digital Preservation,"** *Journal of Research of the National Institute of Standards and Technology*, vol. 107, no. 1, pp. 93-106, 2002.

LEVACOV, Marília. **Bibliotecas virtuais: (r)evolução?**. *Ci. Inf.* [online]. 1997, vol.26, n.2 ISSN 0100-1965. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651997000200003>> Acesso em: 27 de junho de 2012.

MARCHIORI, P. **"Ciberteca" ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação**. *Ciência da Informação*, v. 26, n.2, 1997. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-19651997000200002&script=sci_arttext>. Acesso em: 27 de junho de 2012.

MELLO, Emerson Ribeiro de., Wangham, Michelle S. **Seguranças em Serviços Web**. Minicurso SBSeg 2006.

MONTEIRO, Fernanda., BRÄSCHER, Marisa. **Organização da Informação em Repositórios Temáticos: o uso da modelagem conceitual**. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--261.pdf>>. Acesso em: 29 de junho de 2012.

OLIVEIRA, Renan Rodrigues de. , Carvalho, Cedric Luiz de. **Bibliotecas Digitais e o Repositório Fedora. 2011.** Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-11.pdf>. Acesso em: 01 de junho de 2012.

PHALET, Elena. Préservation Numérique. **Service D'Information Scientifique et Technique, Bruxelles, Union Européenne, juin 2004.** Disponível em: <http://digipat.stis.belspo.be/docs/report_F.pdf>. Acesso em: 27 de junho de 2012.

PRONOM, *The National Archives*. **The Technical Registry PRONOM.** Disponível em: <<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>>. Acesso em: 28 de junho de 2012.

ROSETTO, Marcia. **Os Novos Materiais Bibliográficos e a Gestão da Informação: livro eletrônico e biblioteca eletrônica na América Latina e Caribe.** *Ci. Inf.* [online]. 1997, vol.26, n.1 ISSN 0100-1965. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651997000100008>> Acesso em: 27 de junho de 2012.

SÁ, Werner Martins de. **Biblioteca digital: um estudo sobre a ICDL.** 2009. 107 f. Monografia (Bacharelado em Biblioteconomia)-Universidade de Brasília, Brasília, 2009.

SODRÉ, Muniz. **Apresentação.** Disponível em: <<http://bndigital.bn.br/apresentacao.htm>> Acesso em: 01 de março de 2012.

TAMMARO, Anna Maria; SALARELLI, Alberto. **A biblioteca digital.** Brasília, D.F: Briquet de Lemos, 2008. 365 p.

THOMAZ, Katia P. **Repositórios Digitais Confiáveis e Certificação.** Rio de janeiro, v.3, n.1, p. 80-89, jan./jun.2007

THOMAZ, Katia P. **A preservação digital e o modelo de referência Open Archival Information System (OAIS).** Disponível em: <http://www.dgz.org.br/fev04/Art_01.htm#nota07>. Acesso em: 01 de junho de 2012.

UDDI. *Universal Description, Discovery and Integration.* **Wiki Knowledgebase.** Disponível em: < <http://uddi.xml.org/wiki> > Acesso em: 27 de Junho de 2012.

W3SCHOOLS. **Web Services Basic.** Disponível em: < http://www.w3schools.com/webservices/ws_why.asp > Acesso em: 27 de Junho de 2012.

APÊNDICE A – Código Fonte da Aplicação

Classe Verificar

```

1  package aplicacaows;
2
3  import java.io.File;
4  import java.io.FileInputStream;
5  import java.net.URL;
6  import java.util.logging.Level;
7  import java.util.logging.Logger;
8  import javax.xml.namespace.QName;
9  import org.apache.tika.metadata.Metadata;
10 import org.apache.tika.parser.AutoDetectParser;
11 import org.apache.tika.parser.ParseContext;
12 import org.apache.tika.parser.Parser;
13 import org.apache.tika.sax.BodyContentHandler;
14 import org.xml.sax.ContentHandler;
15
16 /**
17  *
18  * @author LP
19  */
20 public class Verificar {
21
22     public Verificar(){
23         //
24     }
25
26     public MIMEArquivo verificarArquivo(File f) {
27         try {
28             PronomService2 service = new PronomService2(new URL
29                 ("http://www.nationalarchives.gov.uk/pronom/service.asmx?wsdl"),
30                 new QName("http://pronom.nationalarchives.gov.uk", "PronomService"));
31             PronomService port = service.getPronomServiceSoap();
32             SigFile sig = port.getSignatureFileV1();
33
34             String tipo;
35             FileInputStream is;
36             try {
37                 is = new FileInputStream(f);
38
39                 ContentHandler contenthandler = new BodyContentHandler();
40                 Metadata metadata = new Metadata();
41                 metadata.set(Metadata.RESOURCE_NAME_KEY, f.getName());
42                 Parser parser = new AutoDetectParser();
43                 parser.parse(is, contenthandler, metadata, new ParseContext());
44                 tipo = metadata.get(Metadata.CONTENT_TYPE);
45             } catch (Exception e) {
46                 tipo = null;
47                 f = null;
48             }
49
50             for (int i = 0; i < sig.getFFSignatureFile().getFileFormatCollection().getFormat().size(); i++) {
51                 FileFormatType formato = sig.getFFSignatureFile().getFileFormatCollection().getFormat().get(i);
52                 for (int j = 0; j < formato.getInternalSignatureIDOrExtensionOrHasPriorityOverFileFormatID().size(); j++) {
53                     if (tipo.equals(formato.getMIMEType())) {
54                         MIMEArquivo mimeArquivo = new MIMEArquivo();
55                         mimeArquivo.setFileName(f.getName());
56                         mimeArquivo.setMimeType(formato.getMIMEType());
57                         mimeArquivo.setName(formato.getName());
58                         mimeArquivo.setPuid(formato.getPUID());
59                         mimeArquivo.setId(formato.getID().toString());
60                         mimeArquivo.setInternalSignature
61                             (formato.getInternalSignatureIDOrExtensionOrHasPriorityOverFileFormatID()
62                                 .get(j).getValue().toString());
63                         return mimeArquivo;
64                     }
65                 }
66             }
67             return null;
68         } catch (Exception ex) {
69             Logger.getLogger(AplicacaoWS.class.getName()).log(Level.SEVERE, null, ex);
70             return null;
71         }
72     }
73
74 }
75

```

Classe MIMEArquivo

```
1 package aplicacaows;
2
3 /**
4  *
5  * @author LP
6  */
7 public class MIMEArquivo {
8     private String fileName;
9     private String mimeType;
10    private String name;
11    private String puid;
12    private String id;
13    private String internalSignature;
14
15    public MIMEArquivo() {
16        //
17    }
18
19    public MIMEArquivo(String fileName, String mimeType, String name, String puid, String id, String internalSignature) {
20        this.fileName = fileName;
21        this.mimeType = mimeType;
22        this.name = name;
23        this.puid = puid;
24        this.id = id;
25        this.internalSignature = internalSignature;
26    }
27
28    public void setFileName(String fileName) {
29        this.fileName = fileName;
30    }
31
32    public String getFileName() {
33        return fileName;
34    }
35
36    public void setMimeType(String mimeType) {
37        this.mimeType = mimeType;
38    }
39
40    public String getMimeType() {
41        return mimeType;
42    }
43
44    public void setName(String name) {
45        this.name = name;
46    }
47
48    public String getName() {
49        return name;
50    }
51
52    public void setPuid(String puid) {
53        this.puid = puid;
54    }
55
56    public String getPuid() {
57        return puid;
58    }
59
60    public void setId(String id) {
61        this.id = id;
62    }
63
64    public String getId() {
65        return id;
66    }
67
68    public void setInternalSignature(String internalSignature) {
69        this.internalSignature = internalSignature;
70    }
71
72    public String getInternalSignature() {
73        return internalSignature;
74    }
75 }
76
```

Frame jfrmMain

```

1
2 package aplicacaows;
3
4 /**
5  *
6  * @author LP
7  */
8 public class jfrmMain extends javax.swing.JFrame {
9     private Verificar verificar;
10    private MIMEArquivo mimeArquivo;
11    /**
12     * Creates new form jfrmMain
13     */
14    public jfrmMain() {
15        initComponents();
16        verificar = new Verificar();
17    }
18    /**
19     * This method is called from within the constructor to initialize the form.
20     * WARNING: Do NOT modify this code. The content of this method is always
21     * regenerated by the Form Editor.
22     */
23    @SuppressWarnings("unchecked")
24    Generated Code
108
109 private void jmenuAbrirActionPerformed(java.awt.event.ActionEvent evt) {
110     // TODO add your handling code here:
111     this.jtxtAreaTexto.setText("Aguardar...");
112     int returnVal = jFileChooser1.showOpenDialog(this);
113     if (returnVal == jFileChooser1.APPROVE_OPTION) {
114         mimeArquivo = verificar.verificarArquivo(jFileChooser1.getSelectedFile());
115     } else {
116         System.out.println("File access cancelled by user.");
117         this.jtxtAreaTexto.setText("Arquivo Não Selecionado");
118     }
119 }
120
121 this.jtxtAreaTexto.setText("File Name: " + mimeArquivo.getFileName());
122 this.jtxtAreaTexto.setText(this.jtxtAreaTexto.getText() + "\nMIME Type: " + mimeArquivo.getMimeType());
123 this.jtxtAreaTexto.setText(this.jtxtAreaTexto.getText() + "\nName: " + mimeArquivo.getName());
124 this.jtxtAreaTexto.setText(this.jtxtAreaTexto.getText() + "\nPUID: " + mimeArquivo.getPuid());
125 this.jtxtAreaTexto.setText(this.jtxtAreaTexto.getText() + "\nID: " + mimeArquivo.getId());
126 this.jtxtAreaTexto.setText(this.jtxtAreaTexto.getText() + "\nInternal Signature: " +
127     mimeArquivo.getInternalSignature());
128 }
129
130 private void jmenuFecharActionPerformed(java.awt.event.ActionEvent evt) {
131     // TODO add your handling code here:
132     System.exit(0);
133 }
134
135 private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
136     // TODO add your handling code here:
137     jtxtAreaTexto.setText(null);
138 }
139
140 /**
141  * @param args the command line arguments
142  */
143 public static void main(String args[]) {
144     /**
145      * Set the Nimbus look and feel
146      */
147     Look and feel setting code (optional)
148
149     /**
150      * Create and display the form
151      */
152     java.awt.EventQueue.invokeLater(new Runnable() {
153         public void run() {
154             new jfrmMain().setVisible(true);
155         }
156     });
157 }
158
159 // Variables declaration - do not modify
160 private javax.swing.JButton jButton1;
161 private javax.swing.JFileChooser jFileChooser1;
162 private javax.swing.JMenuBar jMenuBar1;
163 private javax.swing.JScrollPane jScrollPane1;
164 private javax.swing.JMenuItem jmenuAbrir;
165 private javax.swing.JMenuItem jmenuArquivo;
166 private javax.swing.JMenuItem jmenuFechar;
167 private javax.swing.JTextArea jtxtAreaTexto;
168 // End of variables declaration
169 }
170

```