colloquium exactarum

# DESCRIÇÃO E RECUPERAÇÃO DE NOTÍCIAS JORNALÍSTICAS POR MEIO DE METADADOS

## DESCRIPTION AND RETRIEVING NEWS STORY BY METADATA

João Brambilla[1], Silvio Carro[1], Munir Felício[1]

[1]Faculdade de Informática – FIPP, Universidade do Oeste Paulista – UNOESTE
E-mail: joaobrambilla@unoeste.edu.br, silvio@unoeste.br, munir@unoeste.br

**RESUMO -** Atualmente existem diversos padrões de metadados para descrição de informações de temas e áreas variadas. Muitos desses padrões estão presentes em sistemas web que visam possibilitar formas de recuperação com maior precisão que os populares mecanismos de busca. O presente trabalho propõe um modelo de metadados para realizar a descrição de notícias e matérias jornalísticas para diversos veículos de informação. Os recursos serão descritos e recuperados por meio de ferramentas computacionais desenvolvidas especificamente para este trabalho.
**Palavras-chave:** Metadados; Xml; Padrões de Metadados.

Recebido em: 08/04/2015
Revisado em: 03/09/2015
Aprovado em: 17/11/2015

**ABSTRACT -** There are currently several metadata patterns which describes information of different themes and areas. Many of these patterns are present in web systems which allow different forms of precisely retrieving than the popular search engines. The current work proposes a metadata model to describe news story in several information vehicles. The resources will be described and retrieved by computational tools which will be developed specifically for this work.
**Keywords:** Metadata; Xml; Metadata Patterns.

## 1 INTRODUÇÃO

A web tem se tornado um veículo cada vez mais utilizado para publicação de notícias, artigos, livros e informações em formato eletrônico em geral, isso devido a seu crescimento desenfreado nos últimos anos (PRAZERES; SANTOS; TEIXEIRA, 2014).

Tendo isto em evidência, a informação vem sendo disponibilizada cada vez mais em formato eletrônico, pois a acessibilidade aumenta com a internet. (PEREIRA; BAPTISTA, 2003), visto que estes recursos estão disponibilizados na rede, se torna imprescindível o desenvolvimento de padrões que visem à descrição dos recursos de informações para posteriormente facilitar a recuperação dessas informações, avaliando sua relevância. Estes padrões são um conjunto de elementos de metadados, que pode ser definido, de maneira simplificada, como dados sobre dados (SOUZA; VENDRUSURO; MELO, 2000).

O desenvolvimento dos metadados deu um grande salto a partir da metade da década de 90, quando foram criados vários modelos de metadados para diferentes tipos de informações. A maneira mais simples de definir um metadado é a de que são dados sobre dados e sua principal função é facilitar a recuperação de informações avaliando sua relevância.

Podemos citar 3 tipos de metadados. O Metadado descritivo, que descreve um recurso para algum fim como descoberta ou identificação. Este tipo pode conter elementos como título, resumo, autor e palavras-chave, Metadado Estrutural, que indica como objetos compostos são unidos, por exemplo, como páginas são ordenadas para formarem capítulos e Metadado Administrativo, o qual fornece informações para ajudar a gerenciar um recurso, por exemplo, como e quando ele foi criado, tipo de arquivo assim como outras informações técnicas e quem pode acessá-lo (NISO, 2004, p. 1).

No contexto do presente trabalho, o metadado do tipo descritivo se adequa perfeitamente para o que se espera como

resultado. Um padrão muito utilizado mundialmente é o Dublin Core (DC, 2013), por possuir um conjunto básico de elementos de descrição e que ainda pode ser estendido para ser mais preciso no que está sendo descrito (SOUZA; VENDRUSURO; MELO, 2000).

Há ainda, outros modelos de metadados descritivos além do Dublin Core, como o MARC (MARC, 2015) e seus desdobramentos e Meta Tags, que também podem ser chamados de modelos intelectuais de metadados. (SILVEIRA, 2014), além dos modelos citados, o modelo que mais tem tido aplicação ultimamente é o RSS (RSS, 2015), acrônimo de RDF Site Summary, Rich Site Summary ou, ainda, Really Simple Syndication.

Segundo Pilgrim (apud ALMEIDA, 2008), "o RSS é um conjunto de especificações voltadas para agregação e distribuição de conteúdo da Web, que facilita o processo de consulta e partilha de informação proveniente de diversas fontes de informação, periodicamente sujeitas a alterações ou atualizações.".

A principal vantagem do RSS é a sua simplicidade, já que o modelo nada mais é do que um arquivo texto codificado em um padrão compatível com o formato XML (eXtensible Markup Language) (ALMEIDA, 2008).

## 1.1 DUBLIN CORE

O Dublin Core é um padrão para descrever recursos de informação. Nos últimos anos, tem sido a principal alternativa como modelo de descrição de metadados por possuir uma utilização simples e trazer consigo os recursos necessários para descrever, identificar, processar, localizar, recuperar e filtrar um documento digital publicado na rede (MODESTO, 2005).

O padrão possui metas e características que, de acordo com Carro (2003), são elas:

- Simplicidade de criação e manutenção. O conjunto de elementos do Dublin Core tem se mantido pequeno e simples, permitindo que pessoas sem experiência possam descrever e submeter seus documentos de maneira facilitada.
- Semântica Comumente Entendida. A busca por informações na Web é caracterizada por uma terminologia e práticas descritivas específicas nas áreas relacionadas. O Dublin Core é uma iniciativa que tenta unificar essas características independente da área da qual a informação é proveniente.
- Extensibilidade. Mesmo priorizando a simplicidade na descrição de recursos digitais, o DC não descarta a necessidade de mecanismos de recuperação de informação mais precisos. Para tanto, existe um esforço no sentido de prover mecanismos de extensão aos elementos DC com o intuito de suportar outras necessidades.

Este padrão de metadados é mantido pela Iniciativa de Metadados Dublin Core (Dublin Core Metadata Initiative).

### 1.1.1 A INICIATIVA DUBLIN CORE

A Iniciative de Metadados Dublin Core (DCMI) mantém inovações compartilhadas em design de metadados e as melhores práticas através de um extenso modelo de propostas e negócios.

Isto é aplicado pelo DCMI da seguinte maneira:

- Gerenciando melhorias de longo prazo e pelo desenvolvimento das especificações e termos de namespaces de metadados;
- Gerenciando uma discussão contínua dos temas de trabalho atuais do DCMI;
- Configurando e gerenciando eventos internacionais e regionais;
- Melhorias e disponibilidade aberta de reuniões incluindo processos, relatórios de projeto e atas de reunião;
- Criação e treinamento nas melhores práticas incluindo tutoriais, conferências web e workshops; e
- Coordenando a comunidade global de voluntários do DCMI.

Os princípios de operação do DCMI são:

- Construção de consenso aberta: A participação na comunidade DCMI é aberta para todos os grupos interessados ou individualmente com experiência ou interesse em metadados. O padrão DCMI de fato mantém, especificações e documentos com as melhores práticas refletidos no consenso que se chegou através de avaliação e debate. Nenhuma taxa é cobrada para utilização de tais informações, uma vez que o valor desses materiais é reforçada por sua ampla adoção.
- Escopo Internacional e participação: A origem do DCMI se deu nos anos 1990 a partir de um workshop informal que atraiu participantes de comunidades mundiais. O DCMI tem sido proposto desde o começo para a participação mundial, como exemplificado por um grande

arranjo de traduções, O local das conferência do Dublin Core e reuniões regionais, e a diversidade de representações regionais entre membros do DCMI e do Comitê de Fiscalização e Conselho Consultivo.

●Imparcialidade de propostas e modelos de negócio: O DCMI é imparcial para propostas de quais padrões de metadados e especificações devem ser usados. O DCMI encoraja a adoção desses padrões e especificações para setores privados e maior padronização de jure que não prejudique o acesso aberto.

●Imparcialidade de tecnologias: O padrão do DCMI é concedido com semânticas (significado de afirmações sobre informações) e recursos. A subjacente infraestrutura tecnológica espera que a codificação e expressão dessas semânticas evoluam com o tempo. O DCMI se preocupa com a independência de manutenção de acordos semânticos, e para a facilidade de expressão dessas semânticas na codificação de idiomas apropriados para a iniciativa de atividade de stakeholders.

●Foco interdisciplinar: Desde sua origem em meados de 90, o princípio de fundação tem sido a descoberta e gerenciamento de recursos através de metadados em suas fronteiras de informações na web e entre redes privadas.

### 1.1.2 ELEMENTOS DO DUBLIN CORE

O conjunto de elementos do padrão de metadados Dublin Core é um vocabulário de quinze propriedades para se usar na descrição de recursos. No começo existiam apenas 13 elementos descritivos no padrão e o nome "Dublin" se dá pela sua origem em 1995 devido a um workshop ocorrido em Dublin, Ohio; "core" devido a seus elementos serem claros e genéricos, podendo ser usados para descrever uma extensa variedade de recursos.

Os 15 elementos do Dublin Core descritos nessa parte são o maior conjunto de vocabulários de metadados e especificações técnicas mantidas pela Iniciativa de Metadados Dublin Core(DCMI).

O conjunto completo de vocabulários, DCMI termos de metadados [DCMI-TYPE], também inclui conjuntos de classes com recursos (Incluindo o tipo de vocabulário do DCMI [DCMI-TYPE]), codificação de esquemas de vocabulário, e codificação de sintaxe de esquemas. Os termos nos vocabulários do DCMI foram projetados para serem usados em combinação com termos de outros vocabulários compatíveis no contexto do perfil da aplicação e na base do modelo abstrato do DCMI [DCAM].

Todas as mudanças feitas para os termos do Conjunto de Elementos Dublin Core desde 2001 foram revisadas por um conselho de uso do DCMI no contexto de uma Política de Namespace [DCMI-NAMESPACE]. A política de namespace descreve como os termos do DCMI são assinados como Identificadores Uniformes de Recursos (URIs) e o limite de conjuntos na escala de medidas editáveis que podem permitir seu uso para rótulos, definições e comentários de uso associados com os termos existentes do DCMI.

A Tabela 1 a apresenta os 15 elementos do Dublin Core.

**Tabela 1.** Os 15 elementos do Dublin Core.

| Element | Elemento |
|---|---|
| 1. Title | 1. Título |
| 2. Creator | 2. Criador |
| 3. Subject | 3. Assunto |
| 4. Description | 4. Descrição |
| 5. Publisher | 5. Editor |
| 6. Contribuitors | 6. Colaborador |
| 7. Date | 7. Data |
| 8. Type | 8. Tipo |
| 9. Format | 9. Formato |
| 10. Identifier | 10. Identificador |
| 11. Source | 11. Fonte |
| 12. Language | 12. Idioma |
| 13. Relation | 13. Relações |
| 14. Coverage | 14. Cobertura |
| 15. Rights | 15. Direitos |

Fonte: (MORATO; MORAES, 2010).

Ainda segundo Morato e Moraes (2010), "Os elementos de metadados Dublin Core podem ser organizados em três grupos (veja Tabela 2) que indicam as características das informações que contemplam."

**Tabela 2.** Grupos com a classificação dos elementos do Dublin Core.

| Relacionados com o conteúdo | Relacionados com a propriedade intelectual do recurso | Relacionados com características formais do recurso |
|---|---|---|
| Title | Creator | Date |
| Subject | Contribuitor | Format |
| Relation | Publisher | Identifier |
| Source | Rights | Language |
| Coverage | | |
| Type | | |

Fonte: (MORATO; MORAES, 2010).

Hoje em dia é fundamental um padrão de metadados para descrever documentos digitais e assim, ter maior precisão no momento de recuperá-las. O Dublin Core é um dos padrões mais utilizados hoje em dia, principalmente, por poder se adequar as necessidades particulares do autor que descreverá o documento, permitindo, assim, criar bases de dados com documentos para áreas específicas.

## 1.2 RSS

A origem do RSS foi dada por uma ideia vinda da Netscape a partir do formato Scripting News em 1999 para que as notícias do site my.netscape.com pudessem ser apresentadas em outros websites. Atualmente, continua sendo utilizado para notícias, mas está mais popular entre os blogs e micro blogs, como o twitter (FERREIRA, 2009).

Ainda de acordo com Ferreira (2009), o significado de RSS foi modificado de acordo com suas versões:

- 0.90, 1.0 – RDF Site Summary
- 0.91, 0.92, 0.93, 0.94 – Rich Site Summary
- 2.0 – Really Simple Syndication

Pilgrim (apud ALMEIDA, 2008), diz que "RSS é um conjunto de especificações voltadas para agregação e distribuição de conteúdo da Web, que facilita o processo de consulta e partilha de informação proveniente de diversas fontes de informação, periodicamente sujeitas a alterações ou atualizações.".

A principal vantagem do RSS é a sua simplicidade, já que o modelo nada mais é do que um arquivo texto codificado em um padrão compatível com o formato XML (eXtensible Markup Language) (ALMEIDA, 2008).

Os documentos disponibilizados pelas empresas ou websites utilizando RSS são comumente chamados de "feeds", que significa "alimentar". Os feeds são listas atualizadas com conteúdo das páginas web (BERNARDINO, 2006).

A identificação das páginas Web que disponibilizam os feeds RSS é caracterizada com um ícone (geralmente na cor laranja) com os acrónimos "RSS" ou "XML". Assim que o fornecedor de informação disponibiliza o feed RSS na respectiva página Web, os utilizadores subscrevem os feeds e procedem à sua leitura através da utilização de programas específicos denominados por agregadores RSS44 ou leitores RSS45. Estes programas agrupam e apresentam os feeds RSS, disponibilizados pelas diversas fontes de informação, permitindo a distribuição dos seus conteúdos facilitando a sua rápida consulta e análise e contribuindo para a partilha de informação".

Ferreira (2009) explica que "Assim como o navegador está para a página HTML, o agregador (reader) está para o arquivo RSS; e da mesma forma que o servidor Web provê página HTML, o publicador (feeder) provê RSS. Um agregador pode ser uma página da Web hospedada num servidor, um programa instalado no computador do usuário final ou um programa instalado no servidor que funciona como um publicador de RSS, fornecendo RSS agregado a outros agregadores."

A Figura 1 ilustra a estrutura de um agregador de conteúdo.

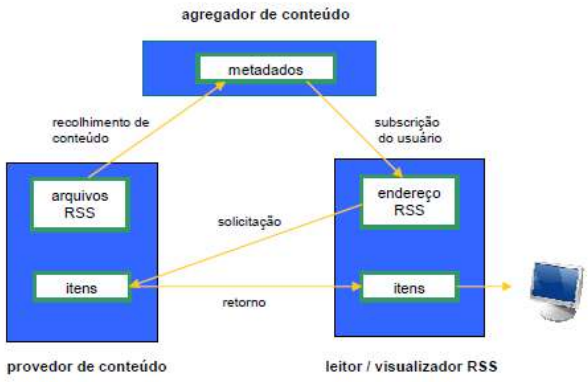



**Figura 1.** Esquema de funcionamento de um sistema baseado em RSS.
Fonte: (ALMEIDA, 2008).

A arquitetura de um feed RSS é escrita na linguagem XML e seus elementos básicos são: Título, descrição e link.

O website About.com discorre com exemplos sobre o código XML gerado para um feed RSS. O website diz que, o RSS usa alguns elementos específicos para distinguir a si mesmo. Uma entrada é a tag <item>. Os elementos para cada página listada nos feeds está dentro da tag <item>. O RSS permite o fornecedor da notícia disponibilizar mais de um conteúdo por vez.

Um leitor ou agregador RSS, deve organizar várias informações em um curto período de tempo. Para isso, estes programas separam os arquivos em canais. Assim como na TV, esta importante tag divide cada feed.

A figura 2 apresenta um exemplo de como seria o XML de um feed RSS seguindo o que foi descrito acima.

Ainda Segundo o site About.com, todo documento XML deve possuir uma declaração de entrada. Adicionalmente, o XML gerado para um feed RSS deve possuir o elemento <rss> como elemento raiz e listar a versão.

```
<channel>
<item>
<title> XML Articles from About.com </title>
<description> Exciting new offerings from the
world of XML and About.com </description>
<link>
http://webdesign.about.com/od/xml-articles-by-
darla-
ferrara/XML_Articles_by_Darla_Ferrara.htm</link
>
</item>
<item>
<title> About.com: Web Design/HTML </title>
<description> Keep up-to-date on all the tips
and tricks in web design with
About.com</description>
<link>http://webdesign.about.com/</link>
</item>
```

**Figura 2.** Exemplo de Feed em XML

A Figura 3 apresenta a versão final do que seria o XML gerado para um feed RSS.

```
<?xml version= "1.0"?>
<rss version= "2.0">
<channel>
<item>
<title> XML Articles from About.com </title>
<description> Exciting new offerings from the
world of XML and About.com </description>
<link>http://webdesign.about.com/od/xml-
articles-by-darla-
ferrara/XML_Articles_by_Darla_Ferrara.htm</link>
</item>
<item>
<title> About.com: Web Design/HTML </title>
<description> Keep up-to-date on all the tips
and tricks in web design with
About.com</description>
<link>http://webdesign.about.com/</link>
</item>
</channel>
</rss></channel>
```

**Figura 3**. Exemplo completo de uma notícia RSS em XML

## 2 METODOLOGIA UTILIZADA

Primeiramente, foi realizado um levantamento sobre os modelos de metadados existentes para selecionar o que mais se adequaria ao trabalho. O modelo Dublin Core, o qual foi identificado como mais adequado em seus elementos para o trabalho, foi selecionado para ser utilizado de maneira a aproveitar seus quinze elementos básicos e estendê-lo, adicionando três novos elementos, que serão: Rank (elemento que dará uma posição à noticia), número da página no jornal e tema da notícia.

Tendo o modelo de metadados definido, os elementos foram estruturados em XML e padronizados via XSD.

Com isso, foi implementado um sistema descritor de notícias utilizando o modelo de metadados selecionado através da tecnologia Java FX. A notícia catalogada é enviada para um Webservice, que também faz parte do projeto em si, desenvolvido através da ferramenta NetBeans 8.0 utilizando a linguagem Java.

Finalizando o projeto, foi implementado o último módulo, que é um sistema web de busca específico para recuperação das notícias catalogadas pelo sistema descritor. Este módulo foi desenvolvido com a utilização da tecnologia JSP/Servlets trazendo três interfaces para o usuário: campos de busca, resultados e conteúdo do resultado.

## 3 METAMODELO NJ

Como discorrido na seção 1.2, o Dublin Core é um padrão projetado para ser usado em combinação com termos de outros vocabulários. Desta forma, se tornou a alternativa mais viável para o projeto.

Neste tópico será apresentada a extensão do Dublin Core para o modelo definido na necessidade de descrever as notícias do Sistema NJ.

### 3.1 ESTENSÃO DUBLIN CORE PARA NJ

O modelo de metadados proposto é representado em UML (Unified Modeling Language) através da Figura 4. Este esquema conceitual, denominado modelo de metamodelo NJ tem o propósito de descrever notícias jornalísticas com relação a impactos ambientais.

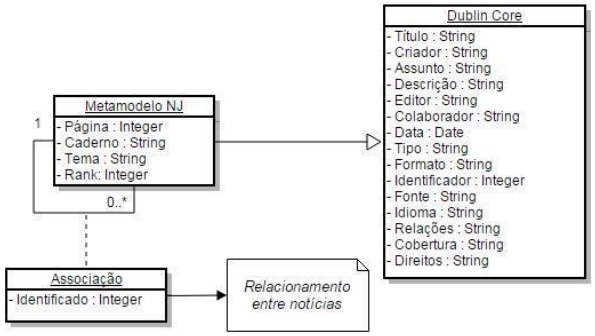



**Figura 4.** Diagrama de classe Metamodelo NJ.

O modelo é fundamentado em quatro novos elementos e um relacionamento entre as notícias: Os elementos dão mais precisão na catalogação de notícias jornalísticas por descrever a página onde a notícia foi publicada, o caderno no jornal, o tema do impacto ambiental e por último o peso que essa notícia obtém identificando o seu grau de importância.

A representação completa do metamodelo NJ é apresentada na Tabela 3.

**Tabela 3.** Metamodelo NJ.

| Elementos | Descrição | Valores/ Formato |
|---|---|---|
| **Título** | Título original da notícia | Texto livre |
| **Criador** | Editor. Pessoa ou organização responsável pelo conteúdo intelectual da notícia | Texto livre |
| **Assunto** | Os tópicos da notícia | Formato: tag1, tag2, tag3 |
| **Descrição** | Descrição textual do recurso, tal como um resumo da notícia. | Texto livre |
| **Editor** | Responsável pela disseminação da notícia no formato atual | Texto livre |
| **Colaborador** | Pessoa ou organização que tenha dado uma contribuição intelectual para a criação da notícia mas num plano secundário | Texto livre |
| **Data** | Data da publicação da notícia no jornal | ISO 8601 aaaa-mm-dd |
| **Tipo** | Tipo ou categoria da notícia | Texto livre |
| **Formato** | Formato ou padrão do recurso | jpg, png, pdf |
| **Identificador** | Valor numérico que não pode ser repetido para representar a notícia | Valor numérico |
| **Fonte** | Nome do jornal onde a notícia foi publicada | Texto livre |
| **Página** | Número da página no jornal onde a notícia foi publicada | Valor numérico |
| **Caderno** | Nome do caderno no jornal onde a notícia foi publicada | Texto livre |
| **Idioma** | Idioma no qual a notícia foi descrita no sistema | ISO 3166 |
| **Relação** | Relacionamento da notícia com outra notícia já cadastrada. O valor inserido no campo diz respeito ao identificador da outra notícia | Valor numérico |
| **Cobertura** | Valor fixo espacial indicando que a notícia é estática | Valor numérico |
| **Direitos** | Direitos autorais da fonte que publicou a notícia | Texto livre |
| **Tema** | Classificação do título em uma determinada temática | Texto livre |
| **Rank** | Importância da notícia | Valor numérico |

## 3.2 DESCRIÇÃO DE NOTÍCIAS UTILIZANDO NJ

A estrutura a seguir representa um exemplo do xml do modelo de metadados proposto. A notícia e seus relacionamentos são descritos utilizando o esquema proposto no modelo de metadados NJ, como apresentado na Figura 5.

```
<metadata
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:nj="http://www.nj.com.br/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" versao="1.07">
        <dc:title>Urutau e filhote ocupam árvore na Sabesp</dc:title>
        <dc:author>João da Silva</dc:author>
        <dc:subject>pássaros; desmatamento; meio ambiente</dc:subject>
        <dc:description>Pássaros, que encontraram morada no estacionamento da
estatal, tentam se camuflar por entre os galhos, esbanjando beleza.
        </dc:description>
        <dc:publisher>José Fagundes</dc:publisher>
        <dc:contribuitor>Organização X</dc:contribuitor>
        <dc:date>2014-11-05</dc:date>
        <dc:type>Ambiental</dc:type>
        <dc:format>PDF</dc:format>
        <dc:identifier>10</dc:identifier>
        <dc:source>
                O Imparcial
                <nj:caderno>Cidades</nj:caderno>
                <nj:pagina>2</nj:pagina>
        </dc:source>
        <dc:language>pt</dc:language>
        <dc:relation>2; 4; 7; 8</dc:relation>
        <nj:tema>Desmatamento</nj:tema>
        <nj:rank>5.5</nj:rank>
        <dc:coverage>false</dc:coverage>
        <dc:rights>licença do autor</dc:rights>
</metadata>
```

**Figura 5.** Modelo de metadados NJ.

## 4 SISTEMA NJ

Com a finalidade de validar o modelo de metadados proposto foi desenvolvido o protótipo de um sistema denominado: Sistema NJ. A arquitetura do sistema NJ é centrada em um serviço Web, que garante o acoplamento leve entre seus subsistemas e através da interação entre eles é possível demonstrar a viabilidade do modelo proposto e a garantia do compartilhamento e a interoperatividade dos metadados.

O sistema NJ é destinado aos profissionais da engenharia ambiental, notícias jornalísticas de impacto ambiental para fins de pesquisa e levantamento de dados em diversos casos ambientais e de vários temas.

O projeto possuía uma questão importante a ser definida: Como armazenar os metadados e o recurso (pdf ou imagem referente à notícia). A solução mais viável encontrada foi o uso de um banco de dados que armazenasse os campos do modelo que fossem pertinentes de busca, o xml na íntegra e o recurso em forma de um array de bytes sendo uma imagem ou um arquivo pdf.

Na Figura 6 a seguir podemos ver o modelo do banco de dados do sistema e a tabela 'noticia' ajustada para solucionar a questão.

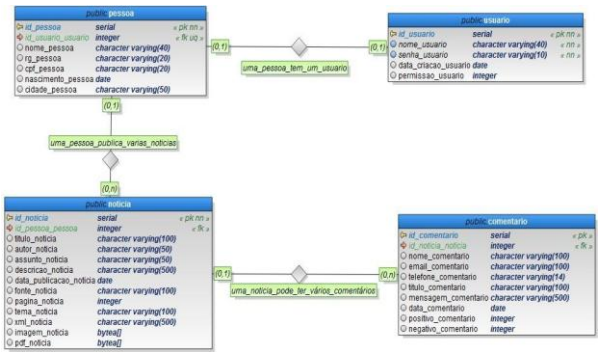



**Figura 6.** Modelo de banco de dados do Sistema NJ.

A questão do compartilhamento das notícias descritas foi resolvida pelo uso da arquitetura cliente-servidor inerente à Internet. O Sistema NJ opera com contribuições voluntárias, portanto a descrição de notícias é distribuída.

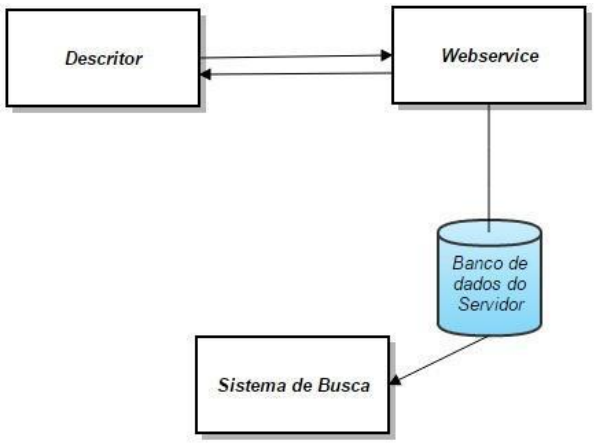



**Figura 7.** Arquitetura do Sistema NJ.

A arquitetura do sistema é composta por três subsistemas: descritor, webservice e sistema de busca (Figura 7). O primeiro subsistema oferece suporte à descrição das notícias baseado no modelo de metadados proposto, insere esta descrição num objeto juntamente com os recursos e os transmite ao subsistema webservice. O webservice mantém uma base de dados com as descrições geradas de cada notícia e seus recursos. A recuperação e visualização são realizadas pelo terceiro subsistema, apoiado por um navegador.

A seguir é detalhado os subsistemas integrantes do sistema NJ.

### 4.1 SISTEMA DESCRITOR




**Figura 8.** Interface do subsistema Descritor.

Desenvolvido através da IDE Netbeans 8.0 com a tecnologia Java FX com a adoção da metodologia cliente-servidor, o subsistema Descritor oferece a interface para a descrição de notícias, importar um recurso, sendo um pdf ou uma imagem, prévia do xml gerado (Figura 8) e o suporte à conexão com o webservice. Dentre os módulos desenvolvidos pode-se destacar: a) interface de descrição de notícias na forma de formulário; b) módulo de importação de recursos nos formatos jpg, png e pdf; c) campo texto com a prévia do xml gerado em tempo real para o usuário; e f) persistência das notícias descritas e

seus recursos no banco de dados do servidor através do webservice.

### 4.2 WEBSERVICE

O webservice foi implementado utilizando a linguagem Java no ambiente de desenvolvimento NetBeans 8.0 através de um sistema web e tecnologia SOAP. Este módulo foi desenvolvido para receber e registrar as notícias e suas dependências no banco de dados do servidor e assim, manter os campos passíveis de busca da notícia preparados para que o módulo de busca possa consultar as notícias na base de dados do servidor.

A seguir, são apresentadas as Figuras 9, 10 e 11 do wsdl gerado referente ao webservice.

```
<definitions xmlns:wsu="http://docs.oasis-open.org/wss/2004/01/oasis-200401-wss-wssecurity-utility-1.0.xsd"
xmlns:wsp="http://www.w3.org/ns/ws-policy" xmlns:wsp1_2="http://schemas.xmlsoap.org/ws/2004/09/policy"
xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata" xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:tns="http://webservice/" xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns="http://schemas.xmlsoap.org/wsdl/"
targetNamespace="http://webservice/"
name="WebServiceNj">
    <types>
        <xsd:schema>
        <xsd:import namespace="http://webservice/" schemaLocation="http://localhost:8084/SistemaBuscaNj/WebServiceNj?xsd=1"/>
        </xsd:schema>
    </types>
    <message name="recuperaNoticias">
        <part name="parameters" element="tns:recuperaNoticias"/>
    </message>
    <message name="recuperaNoticiasResponse">
        <part name="parameters" element="tns:recuperaNoticiasResponse"/>
    </message>
    <message name="enviarParametros">
        <part name="parameters" element="tns:enviarParametros"/>
    </message>
    <message name="enviarParametrosResponse">
        <part name="parameters" element="tns:enviarParametrosResponse"/>
    </message>
    <message name="autenticarUsuario">
        <part name="parameters" element="tns:autenticarUsuario"/>
    </message>
    <message name="autenticarUsuarioResponse">
        <part name="parameters" element="tns:autenticarUsuarioResponse"/>
    </message>
```

**Figura 9.** Primeira imagem referente ao wsdl do webservice NJ.

```
<portType name="WebServiceNj">
    <operation name="recuperaNoticias">
        <input wsam:Action="http://webservice/WebServiceNj/recuperaNoticiasRequest" message="tns:recuperaNoticias"/>
        <output wsam:Action="http://webservice/WebServiceNj/recuperaNoticiasResponse" message="tns:recuperaNoticiasResponse"/>
    </operation>
    <operation name="enviarParametros">
        <input wsam:Action="http://webservice/WebServiceNj/enviarParametrosRequest" message="tns:enviarParametros"/>
        <output wsam:Action="http://webservice/WebServiceNj/enviarParametrosResponse" message="tns:enviarParametrosResponse"/>
    </operation>
    <operation name="autenticarUsuario">
        <input wsam:Action="http://webservice/WebServiceNj/enviarParametrosRequest" message="tns:autenticarUsuario"/>
        <output wsam:Action="http://webservice/WebServiceNj/enviarParametrosResponse" message="tns:autenticarUsuarioResponse"/>
    </operation>
</portType>
```

**Figura 10.** Segunda imagem referente ao wsdl do webservice NJ.

```
    <binding name="WebServiceNjPortBinding" type="tns:WebServiceNj">
        <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="document"/>
        <operation name="recuperaNoticias">
            <soap:operation soapAction=""/>
            <input>
                <soap:body use="literal"/>
            </input>
            <output>
                <soap:body use="literal"/>
            </output>
        </operation>
        <operation name="enviarParametros">
            <soap:operation soapAction=""/>
            <input>
                <soap:body use="literal"/>
            </input>
            <output>
                <soap:body use="literal"/>
            </output>
        </operation>
        <operation name="autenticarUsuario">
            <soap:operation soapAction=""/>
            <input>
                <soap:body use="literal"/>
            </input>
            <output>
                <soap:body use="literal"/>
            </output>
        </operation>
    </binding>
    <service name="WebServiceNj">
        <port name="WebServiceNjPort" binding="tns:WebServiceNjPortBinding">
            <soap:address location="http://localhost:8084/SistemaBuscaNj/WebServiceNj"/>
        </port>
    </service>
</definitions>
```

**Figura 11.** Terceira imagem referente ao wsdl do webservice NJ.

Segundo mostra Sanchez (2011), WSDL é um documento proposto pela W3C a partir de Junho de 2007 escrito em XML que visa padronizar as descrições das funcionalidades oferecidas por web services de forma independente de plataforma ou linguagem. Possui basicamente duas finalidades:

1. Expor os métodos que determinado serviço disponibilizará
2. Possibilitar a localização de determinado serviço.

## 4.3 SISTEMA DE BUSCA

O Visualizador NJ é baseado na geração de páginas em HTML. As páginas são responsáveis pela seleção dos atributos de busca, pela apresentação e o detalhamento dos resultados (Figuras 12, 13 e 14). As páginas são geradas através da linguagem Java com tecnologia jsp e servlet.

O subsistema oferece dois tipos de consulta: simplificada e avançada. Os resultados da busca são apresentados em outra página com os detalhes da notícia, possibilidade de comentário, feedback do usuário e os recursos disponíveis para download.

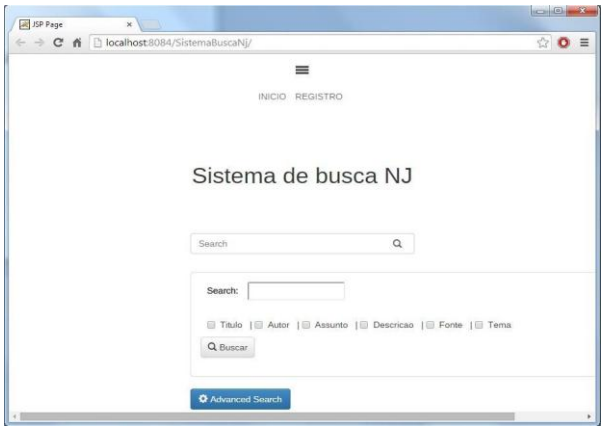



**Figura 12.** Módulo de busca do sistema NJ.




**Figura 13.** Módulo de exibição dos resultados do sistema de busca NJ.

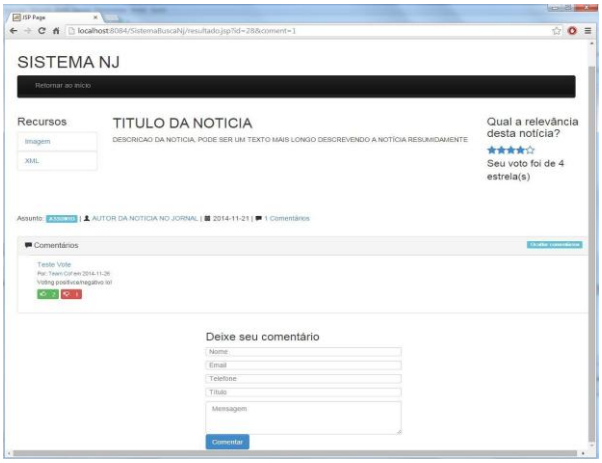



**Figura 14.** Módulo de apresentação do resultado do sistema de busca NJ.

## 5 TESTES E RESULTADOS

Para a realização de testes iniciais, foram inseridas três notícias extraídas de web sites com o tema sobre meio ambiente no sistema NJ. O processo de pesquisa foi por navegação.

Em um primeiro momento foi realizada a busca pelas notícias cadastradas por meio do módulo avançado do NJ.

Posteriormente, a mesma consulta foi realizada em sistemas de buscas tradicionais da Web e um sistema específico, analisando a primeira página de resultados: Google, Bing e AchaNoticias.

**Tabela 4.** Resultado da busca em sistemas web.

| Sistemas | Resultados | Resultados dentro do contexto | Resultados Precisos | Precisão (%) |
|---|---|---|---|---|
| Sistema NJ | 3 | 3 | 2 | 66% |
| Google | 6 | 2 | 1 | 16,6% |
| Bing | 10 | 5 | 0 | 0% |
| AchaNotícias | 10 | 3 | 2 | 20% |

## 6 CONSIDERAÇÕES FINAIS

Este trabalho propõe um modelo de metadados para compartilhar notícias jornalísticas de caráter ambiental na Web e apresenta um protótipo baseado neste modelo para a recuperação e visualização das notícias.

A motivação principal do modelo de metadados e do sistema NJ foi desenvolver uma solução que permita o acesso aos casos publicados de notícias, ora confinadas nas mais diversas instituições de notícias e nos mais diversos meios. A ideia de usar arquitetura de metadados do Dublin Core para a formatação dos dados e a de disponibilizar esses dados juntamente com o modelo permite que os recursos gerados possam ser utilizados e intercambiados com outros sistemas. A independência existente entre o modelo de metadados e o padrão das notícias torna facilmente possível a extensão para outros padrões.

É pretendido em um futuro próximo, testar o sistema NJ na área acadêmica disponibilizando o conteúdo para universitários, mestrandos e doutorandos, habilitando uma nova ferramenta para as pesquisas de seus respectivos projetos de pesquisa.

Neste contexto, o sistema NJ é um importante recurso para a rápida distribuição do conhecimento e da informação na área de notícias com caráter ambiental.

## REFERÊNCIAS


ALMEIDA, R. L. **Disseminação de conteúdo na web:** a tecnologia RSS como proposta a comunicação científica. Brasília, 2008. Disponível em: <http://eprints.rclis.org/11856/1/disserta%C3%A7%C3%A3o_final_rss.pdf>. Acesso em: 20 set. 2013.

BERNARDINO, T. S. M. P. **Perspectiva sobre a utilização da tecnologia RSS no contexto da comunicação científica**. 2006. Dissertação (Mestrado em Sistemas de Informação) - Universidade do Minho.

CARRO, S. A. **Um modelo de metadados para a indexação e recuperação de imagens médicas na web**. 2003. Dissertação (Mestrado em Ciência da Computação) - Porto Alegre - RS.

DC. **Dublin Core metadata initiative.** Disponível em: <http://dublincore.org/documents/dces/#DCTERMS>. Acesso em: 11 out. 2013.

FERREIRA, M. C. **iRSS – uma ferramenta de agregação de RSS baseada em taxonomia**. 2009. Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro – RJ.

MARC STANDARDS**.** Disponível em: <http://www.loc.gov/marc/>. Acesso em: 19 nov. 2015.

MODESTO, F. **Metadados:** introdução básica. 2. ed. São Paulo: Universidade de São Paulo, 2005. Disponível em: <http://www.eca.usp.br/prof/fmodesto/textos/livrometadados.pdf>. Acesso em: 11 out. 2013.

MORATO, A. C; MORAES, M. A. **Metadados, Dublin Core:** uma breve introdução. 2010. Disponível em: <http://eprints.rclis.org/14424/1/Dublin_Core_-_uma_breve_introdu%C3%A7%C3%A3o.pdf. Acesso em: 11 out. 2013.

NISO. **Understanding Metadata.** Bethesda, MD: NISO Press, 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Acesso em: 20 set. 2013.

PRAZERES, C. V. S; SANTOS, C. A. S; TEIXEIRA, C. A. C. **Produção de Notícias Multimídia para**

**Diferentes Meios de Comunicação.** Salvador: Universidade de Salvador. Disponível em: <http://homes.dcc.ufba.br/~prazeres/papers/NewsProd.pdf> Acessado em: 14 ago. 2014.

PEREIRA, T.; BAPTISTA, A. A. **Omnipaper:** descrição de recursos de notícias digitais em RDF**.** p.2, 2003.

RSS ADVISORY BOARD**.** Disponível em: <http://www.rssboard.org/rss-specification>. Acesso em: 19 nov. 2015.

SANCHES, F. **WSDL - O que é? Pra que serve? Onde utilizo?**, 8 set. 2011. Disponível em: <http://fabriciosanchez.com.br/2/wsdl-o-que-e-pra-que-serve-onde-utilizo>. Acesso em: 12 dez. 2014.

SILVEIRA, M. **METADADOS:** histórico, conceitos, aplicações e formatos. Universidade Federal de Pernambuco, Disponível em: <http://www.slideshare.net/lulinha/metadados>. Acessado em: 14 ago. 2014.

SOUZA, M. I. F.; VENDRUSURO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, p. 93-102, jan./abril 2000.