

Padrão Dublin Core de Metadados e Microformatos Dublin Core como Tecnologias de Descrição de Informações na Web Semântica

Ms. Eveline Batista Rodrigues
Centro Universitário Eurípides de Marília – São Paulo – Brasil
evelinerodrigues@yahoo.com

Dr. José Eduardo Santarém Segundo
Centro Universitário Eurípides de Marília – São Paulo – Brasil
Faculdade de Tecnologia de Lins - Fatec Lins - Brasil
santarem@univem.edu.br

Resumo – Este trabalho tem o objetivo de apresentar as potencialidades do padrão de metadados Dublin Core e os microformatos Dublin Core cujo objetivo é a descrever as informações na web semântica. Serão apresentadas soluções práticas para a catalogação dos dados e descrição dos recursos eletrônicos, mostrando como é possível descrever, através desses padrões, informações com significado, de modo a tornar as coleções de dados mais visíveis pelos engenhos de busca e sistemas de recuperação.

Palavras-chave: Dublin Core, metadados, Microformatos Dublin Core, Web Semântica.

Abstract – This paper aims to present the potential of Dublin Core metadata standard and Dublin Core microformats whose goal is to describe the information in the semantic web. Will be presented practical solutions to the cataloging of data and description of electronic resources, showing how you can describe, through these patterns, information with meaning, to make the data collections more visible to search engines and retrieval systems.

Keywords: Dublin Core, metadata, Dublin Core Microformats, Semantic Web.

Introdução

No fim do Século XX, surge a Sociedade da Informação, assinalando um desenvolvimento social e econômico onde a informação, como meio de criação de conhecimento, desempenha um papel fundamental na produção de riqueza e na contribuição para o bem-estar e qualidade de vida das pessoas. Uma das premissas para o avanço deste tipo de sociedade é a possibilidade das pessoas de acederem às Tecnologias de Informação e Comunicação.

Este tipo de sociedade colocou à disposição de milhões de internautas uma gama muito grande de informações, das quais o deixou muitas vezes desorientado e com dificuldade de procurar o que desejava.

Para solucionar o problema do excesso de informação e à informação que se desejava obter, foram criadas as ferramentas de busca, como as das

empresas Altavista, Yahoo e Google, que faziam a filtragem da informação através de meta-informações, como palavras-chave, além de heurísticas para categorização [1].

Os robôs de busca são softwares responsáveis pela busca e localização dos recursos informacionais, vasculhando a *Web* utilizando estratégias variadas para se locomoverem de um site a outro, geralmente começando a busca pelos sites mais populares e a partir da *homepage*, vão seguindo os *links* e adicionando os endereços ao banco de dados. Usam algoritmos próprios para determinar os links a serem seguidos e voltam aos sites regularmente para verificar as alterações e atualizar o sistema.

Atualmente, os motores de busca já se mostram ineficientes, devido à quantidade de informação indesejada que o usuário comum recebe ao fazer uma simples busca. Isso acontece devido a vários problemas, como: 1) baixa qualidade da indexação, por ser feita automaticamente, que resulta em grande quantidade de informações recuperadas, a maioria sem relevância; 2) cobertura parcial da Internet; 3) as ferramentas de busca não são especializadas; 4) indexam páginas HTML isoladas e não recursos; 5) grandes quantidades de informações disponíveis na Internet estão sob a forma de registros contidos em bases de dados, acessados somente por meio das interfaces destas bases de dados, e desta forma ficam inacessíveis aos programas robôs [2].

Da evolução de uma Web Sintática, cheia de regras e formalidades, nasce o projeto da Web Semântica, que consiste não apenas em um grande livro hiperligado onde pessoas possam pesquisar, navegar, ler e visualizar informações, mas possibilita que computadores interpretem, estabeleçam inferências e relações na web, compartilhando informações (interpretações) com organização em escala e completa integração de recursos [3].

A Web Semântica, tendo como precursor Tim Berners-Lee *et. al.*, tem sido indicada como um caminho para solucionar a representação dos recursos informacionais na Web. Ela é um aprimoramento da Web atual, objetivando classificar e dar semântica a todo o conteúdo disponível na rede, a fim de proporcionar um melhor tratamento e recuperação das informações nela contidas, de modo que seus usuários possam buscar, localizar e recuperar facilmente e eficientemente o que procuram [4].

Desta forma, Berners-Lee [4] imagina um mundo em que programas e dispositivos especializados e personalizados, chamados agentes, possam interagir por meio da infra-estrutura de dados da Internet trocando informações entre si, de forma a automatizar tarefas rotineiras dos usuários. Estabeleça uma língua franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral.

O estabelecimento da Web Semântica se faz através do trabalho conjunto de várias outras ferramentas tecnológicas, entretanto a base para sua construção está no uso de metadados para a representação dos recursos informacionais.

Assim, uma das maneiras de dar semântica às informações, para que possam ser pesquisadas e recuperadas com sentido pelas pessoas – são os metadados – que são dados sobre os dados. É a catalogação do dado ou descrição do recurso eletrônico.

Desta forma o trabalho justifica-se pela necessidade de conhecer novas tecnologias que possam tornar as coleções de informações mais visíveis pelos engenhos de busca e sistemas de recuperação, e que esta informação recuperada possa ser dotada de significado para o usuário.

Metodologia

Metadados e o Padrão Dublin Core

Os metadados são “dados sobre dados”, o que quer dizer que metadados referem-se a estrutura descritiva da informação sobre outro dado, o qual é usado para ajudar na identificação, descrição, localização e gerenciamento de recursos da web. Entretanto, eles podem ser aplicados em qualquer meio [5].

O padrão de metadados Dublin Core (DC) é uma iniciativa para criação de um vocabulário controlado, mesmo que limitado - para uso na Web, baseado no pressuposto de que a busca por recursos de informação deve ser independente do meio em que estão armazenadas [6].

O Dublin Core Metadata Initiative (DCMI) ganhou o nome da localidade onde se deu o encontro inicial, Dublin, Ohio, USA [7]. É um projeto destinado a organizar as informações nas páginas da WEB, com o objetivo de estabelecer padrões de catalogação e classificação das informações no meio eletrônico. Sua aceitação foi rápida e é hoje um padrão internacional, com participantes de mais de 20 países.

No escopo do projeto da Web Semântica surgem os padrões de metadados, de construção dos códigos XML. Estabelecido pelo Consórcio W3C, responsável pelo gerenciamento da Internet, propicia um conjunto 15 elementos de metadados [7].

Seus elementos são: **DC.Title** [Título], **DC.Creator** [Autor / Criador], **DC.Subject** [Assunto], **DC.Description** [Descrição], **DC.Publisher** [Editor ou Produtor], **DC.Other Contributors** [Colaboradores], **DC.Date** [Data], **DC.Type** [Gênero], **DC.Format** [Formato], **DC.Identifier** [Identificação], **DC.Source** [Fonte], **DC.Language** [Idioma], **DC.Relation** [Relações], **DC.Right** [Direito Autoral]e **DC.Coverage** [Cobertura].

Aplicação de Metadados Dublin Core

O objetivo da aplicação é descrever através de metadados Dublin Core o conteúdo do site **www.estamosnaweb.com**, cujo conteúdo é o vídeo oficial da Copa do Mundo 2010, conforme a figura 1:

Trailer da Copa do Mundo 2010

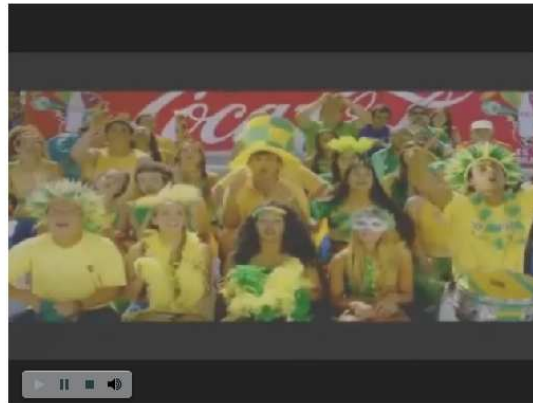


Figura 1. Conteúdo do site estamos na web.com.

A tag META insere as metainformações no cabeçalho do documento. Abaixo (figura 2) os metadados DC inseridos dentro da tag **<head>** do HTML:

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />

<title>Video Oficial da Copa do Mundo 2010</title>

<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
  <meta name="DC.title" content="Video oficial da Copa do Mundo 2010"/>
  <meta name="DC.identifier" content="www.estamosnaweb.com"/>
  <meta name="DC.description" content="Video oficial da Copa do Mundo na Africa do Sul 2010"/>
  <meta name="DC.subject" content="Video Oficial, Copa do Mundo 2010, Africa do Sul"/>
  <meta name="DC.language" scheme="ISO639-1" content="pt"/>
  <meta name="DC.creator" content="www.fifa.com"/>
  <meta name="DC.publisher" content="www.estamosnaweb.com"/>
  <meta name="DC.type" scheme="DCMITYPE" content="http://purl.org/dc/dcmitype/MovingImage"/>
  <meta name="Dc.date" content="2010" />
  <meta name="Dc.type" content="video" />
  <meta name="Dc.format" content="flv" />
  <meta name="classification" content="Copa 2010, Música, Video, Africa do Sul"/>
</head>
```

Figura 2. Metadados DC para descrição do vídeo oficial da copa do mundo 2010.

Microformatos Dublin Core

Microformatos são porções de código HTML ou XHTML cujo objetivo é adicionar conteúdo semântico aproveitando as características dos atributos "id" ou "class" usado por algumas tags HTML. Desta forma, atribuímos um valor aos atributos "id" ou "class", que além de associar um estilo de apresentação através de folhas de estilo (por exemplo, cor, tamanho, tipo de fonte,...) também possuem outra utilidade. [8]

O principal objetivo dos microformatos, é que os metadados sejam úteis para as pessoas em primeiro lugar, e depois aos agentes de usuário (por exemplo, buscadores).

Microformatos são um conjunto de formatos de dados abertos e simples construídos sobre padrões existentes e amplamente aceitos. Pretendem resolver problemas simples primeiramente adaptando a comportamentos e padrões atuais de uso (ex. XHTML) [9].

Microformatos são:

- Uma maneira de pensar sobre dados;
- Projeto de princípios para formatos;
- Adaptado de comportamentos e padrões atuais de uso;
- Altamente correlacionado com a semântica XHTML;
- Um conjunto de padrões abertos simples formato de dados que muitos estão ativamente desenvolvendo e implementando para uma melhor estruturação de blogs e publicação micro conteúdo web em geral;
- "Uma revolução evolutiva".

Através dos Microformatos, as páginas podem trocar informações e permitir também que agentes recuperem essa informação de forma mais clara e ágil do que quando esse tipo de informação fica disponível apenas em texto puro, em formato HTML.

Os pesquisadores Eva Méndez, Alejandro Bravo e Leandro Mariano López criaram um novo conceito para embutir informações sobre um recurso na Web, através do uso de dois padrões consolidados: Dublin Core, para descrição de recursos digitais na Web, e Microformatos, para embutir metainformações em ambientes web.

Eles juntaram todos os elementos de metadados DC na lista de Microformatos, e dessa forma permitem reforçar as indiscutíveis vantagens do padrão DC (simplicidade, flexibilidade e adequação para qualquer domínio) para descrever, através dos Microformatos DC, qualquer recurso que se deseja citar em um documento XHTML [10].

A principal função destas especificações é enriquecer a informação inserida em páginas web com metainformação, e isso é feito codificando os metadados no corpo do documento. O nome Microformatos está relacionado com a informação de "pequenos formatos" (micro + format) de dados (informação) válidos no código do seu conteúdo XHTML. Os Microformatos podem ocupar o lugar que antes era ocupado pela *tags* META do HTML. As *tags* META tinham o objetivo de apresentar metainformações sobre o conteúdo da página, sendo inclusive utilizados pelas ferramentas de busca para compor o banco de dados e, conseqüentemente, servindo de base para constituição do valor a ser recuperado [11].

Aplicação de Microformatos Dublin Core

Da mesma forma, para aplicarmos os Microformatos Dublin Core usaremos o conteúdo do site www.estamosnaweb.com, cujo conteúdo é o vídeo oficial da Copa do Mundo 2010.

A diferença entre as *tags* META e os Microformatos é que, neste segundo, as informações são cadastradas no corpo do documento, diferente das *tags* META, que inseriam as metainformações no cabeçalho do documento.

Observe que os microformatos são colocados dentro da *tag* **<body>** do HTML, conforme a figura 3.

```

<dl class="dublincore">
  <dt><b>Titulo:</b></dt>
  <dd class="title">Video oficial da Copa do Mundo 2010</dd>
  <dt><b>Identificador (URL):</b></dt>
  <dd><a href="www.estamosnaweb.com" class="identifier">www.estamosnaweb.com</a></dd>
  <dt><b>Descrição:</b></dt>
  <dd class="description">Video oficial da Copa do Mundo na Africa do Sul 2010</dd>
  <dt><b>Palavras chave:</b></dt>
  <dd class="subject">Video Oficial, Copa do Mundo 2010, Africa do Sul</dd>
  <dt><b>Codigo de idioma:</b></dt>
  <dd class="language">pt</dd>
  <dt><b>Autor:</b></dt>
  <dd><a href="www.fifa.com" class="creator">Fifa</a></dd>
  <dt><b>Editora:</b></dt>
  <dd><a href="www.estamosnaweb.com" class="publisher">Eveline Batista Rodrigues</a></dd>
</dl>

```

Figura 3. Utilização de Microformatos DC para descrição do vídeo oficial da copa do mundo 2010.

Na primeira linha de código (`<dl class="dublincore">`) temos um elemento HTML, neste caso, uma lista de definição **dl**. Este elemento tem uma classe `dublincore`. Esta classe pode ter uma folha de estilo de apresentação: fonte, cor, tamanho, bordas, e assim por diante. Mas, também pode ter um conteúdo semântico. Com a classe `dublincore` pretendemos indicar aos agentes de usuário (buscadores) algo como: *Atenção, dentro deste elemento você vai encontrar elementos definidos pelo Dublin Core, lembre-se de obter informação semântica.*

Na linha seguinte, encontramos `<dt>Titulo</dt>`, ou seja, "conforme definido" em uma lista de definições: neste caso o título. Qualquer pessoa que obtiver este código no seu navegador verá escrita a palavra "Titulo" e entenderá que abaixo aparecerá o título do conteúdo.

Depois encontramos o outro elemento necessário para uma boa lista de definição, `<dd>`, nomeadamente a definição. Neste caso, encontramos um texto.

Abaixo encontramos uma ligação, com uma característica especial: **class = "title"**. Essa classe tem um nome que corresponde a um dos elementos Dublin Core e os agentes de usuário podem ver que este rótulo é o título. Além disso, como está dentro de um elemento de uma classe `dublincore`, se um agente de usuário conhecer os microformatos Dublin Core poderia associar facilmente o **title** (o título) usado no microformato com o termo **title** definido pelo Dublin Core.

Resultados

Podemos checar os metadados Dublin Core através de um *add-on* disponível para o navegador Mozilla Firefox, chamado de Dublin Core Viewer 0.8, no qual mostra os metadados Dublin Core incorporados em documentos HTML/XHTML com elementos META e LINK.

A figura 4 mostra o acesso à ferramenta pelo navegador Mozilla Firefox, através de um pequeno símbolo no canto inferior direito, para que através dele seja apresentado o quadro com as informações do recurso.

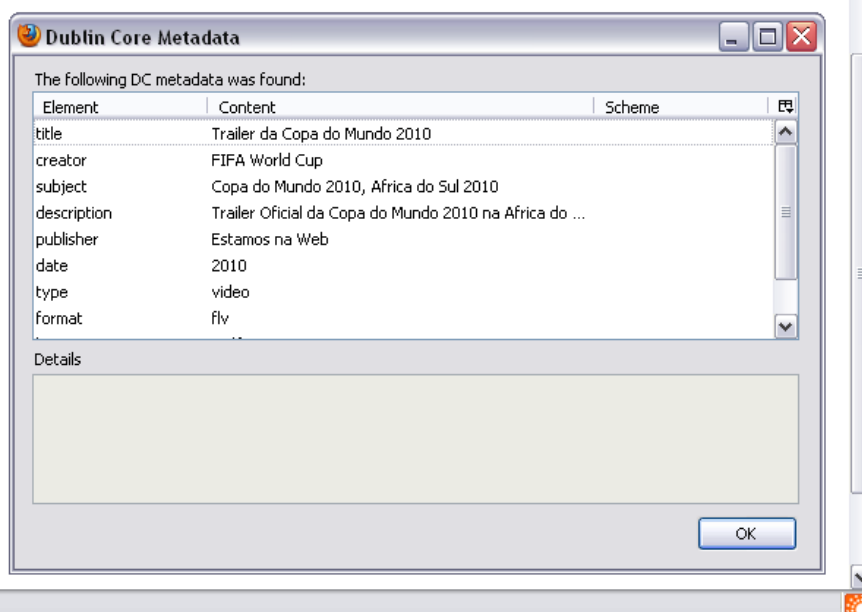


Figura 4. Janela do Dublin Core Viewer Extension exibindo a descrição do recurso com os metadados DC.

Os microformatos descritos na Figura 3 são exibidos quando é exibida a página HTML onde são utilizados, pois são colocados dentro da *tag <body>*, pois o objetivo dos microformatos é apresentar estas informações aos humanos.

A figura 5 mostra como o usuário visualizará os microformatos no site estamosnaweb.com:

Trailer da Copa do Mundo 2010



Título:
Video oficial da Copa do Mundo 2010

Identificador (URL):
www.estamosnaweb.com

Descrição:
Video oficial da Copa do Mundo na Africa do Sul 2010

Palavras chave:
Video Oficial, Copa do Mundo 2010, Africa do Sul

Código de idioma:
pt

Autor:
[Fifa](#)

Editora:
[Eveline Batista Rodrigues](#)

Figura 5. Visualização do site estamosnaweb.com com as metainformações descritas pelos Microformatos Dublin Core.

Apesar dos Microformatos já estarem disponíveis há algum tempo, ainda são poucas as ferramentas e agentes que exploram as páginas que contêm essas informações embutidas. Alguns browsers, como o Mozilla Firefox e o Opera, através de *add-ons* e *extensions*, deixam disponível ao usuário, se este desejar, a inclusão de agentes que detectam e permitem interatividade através dos Microformatos.

Existe um *add-on* disponível para o navegador Mozilla Firefox, chamado de Dublin Core Microformats Viewer 2.1.6, que mostra os microformatos Dublin Core existentes na página. A figura 6 mostra o acesso à ferramenta pelo navegador Mozilla Firefox, através de um pequeno símbolo no canto inferior direito, para que através dele sejam apresentados os microformatos encontrados na página.

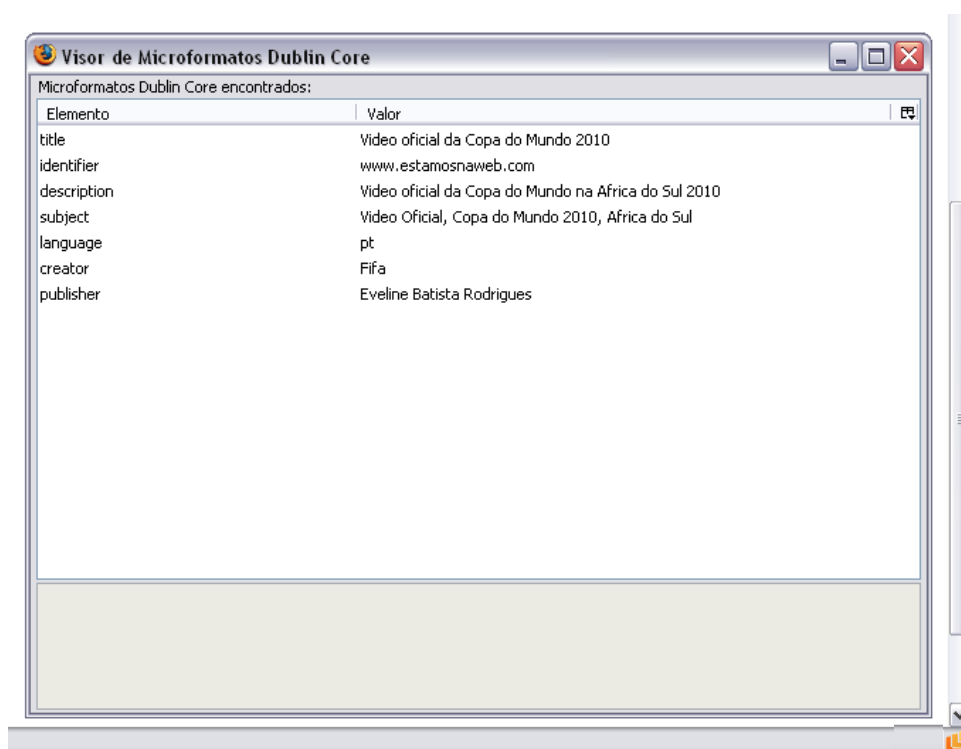


Figura 6. Janela do Dublin Core Microformats Viewer exibindo os microformatos DC encontrados no site estamosnaweb.com.

A criação de metadados e a utilização do Microformato Dublin Core podem ser auxiliadas através de uma ferramenta online disponível para criação e uso da especificação. Esta ferramenta é o “Dublin Core Metadata Gen”, um gerador online de metadados e microformatos Dublin Core (<http://webpossible.com/utilidades/dublincore-metadata-gen/index.php?lang=en>) [12].

Discussão e Conclusões

Com o surgimento e efetivação da Web Semântica como um caminho a ser seguido, existe claramente uma preocupação na estruturação da informação, principalmente no que diz respeito à separação de

estrutura e conteúdo na geração de informações da Web. Esse contexto fornece as condições necessárias para que a recuperação semântica da informação seja possível, completando o ciclo de técnicas e métodos que compõe a Web Semântica.

No que tange a organização e estruturação da informação na web, o uso de um formato de metadados como elemento norteador tem sido fundamental para que o contexto de armazenamento seja compatível com a proposta da Web Semântica, com destaque para o formato Dublin Core, dado sua estrutura funcional que permite interoperabilidade, extensibilidade e principalmente sua aceitação e posicionamento com forte uso dentre os modelos de ambientes integradores de informação digital.

A utilização de microformatos embutidos em páginas web, com destaque aos microformatos DC, é uma idéia inicial do que se pode ser feito em relação a organização e recuperação de informações, sem que haja interferência humana, de forma que a troca de informações permita aos próprios browsers e sistemas para web identificar e trocar informações de forma semântica.

Fica claro que a tecnologia apresentada neste trabalho não efetiva a Web Semântica, porém faz parte de um conjunto de peças, que ainda contém outras como ontologias e técnicas de recuperação da informação, que juntas caminham para que em breve seja possível a recuperação da informação com capacidade semântica.

Referências

- [1] BRANDÃO, A. A. F.; LUCENA, C. J. P. **Uma Introdução à Engenharia de Ontologias no contexto da Web Semântica**. Departamento de Informática PUC do Rio de Janeiro. Rio de Janeiro. PUC-RioInf. MCC29/02 Novembro, 2002
- [2] MARCONDES, C. H.; SAYÃO, L. F. **Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira**. Ciência da Informação, Brasília.dd2001. Disponível em: < http://www.scielo.br/scielo.php?pid=S0100-19652001000300004&script=sci_arttext&tIng=es>. Acesso em: 09 dez. 2009.
- [3] DINIZ, V.; CECCONI, C. **O W3C e a Web Semântica**. WorkShop Rede IP do Futuro. Abril de 2009.
- [4] BERNERS-LEE, T.; LASSILA, O.; HENDLER, J. **The semantic web**. Scientific American, 284(5), 34-43. Maio, 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>. Acesso em: 28 nov. 2009.
- [5] DZIEKANIAK, G. V.; KIRINUS J. B. **Web Semântica**. Revista Eletrônica de Biblioteconomia e Ciência da Informação. 2004. Volume:9 Issue:18 pages/rec. No:20-39. Disponível em:<<http://www.doaj.org/doaj?func=abstract&id=190138>>. Acesso em: 09 dez. 2009.

- [6] SOUZA, R. R.; ALVARENGA, L. **A Web Semântica e suas contribuições para a ciência da informação.** Revista Ci. Inf. vol.33 no.1 Brasília Jan./Apr. 2004.
- [7] DUBLIN CORE METADATA INITIATIVE. Disponível em: <<http://dublincore.org>>. Acesso em: 09 jun. 2010.
- [8] GARCIA. Alejandro G. B. **Introducción a los microformatos.** WebPosible. Disponível em: <<http://www.webposible.com/microformatos-dublincore/introduccion-microformatos.html>>. Acesso em 29 jun. 2010.
- [9] MICROFORMATS.ORG. **About Microformats.** Disponível em: <http://microformats.org/about>. Acesso em: 29 jun. 2010.
- [10] MÉNDEZ, E.; BRAVO, A.; LÓPEZ, L. M. **Microformatos: web 2.0 para Dublin Core.** El profesional de la información, Barcelona, v. 16, n. 2, p. 107-113, marzo/abr. 2007.
- [11] SEGUNDO, José Eduardo S. **Representação Iterativa. Um Modelo para Repositórios Digitais.** Tese de Doutorado. Faculdade de Filosofia e Ciências. UNESP. Campus de Marília/SP. 2010.
- [12] WEBPOSIBLE 2010. **Microformats Dublin Core in a few (and easy) steps.** Disponível em: <http://webposible.com/microformatos-dublincore/microformats_dublin-core.html>. Acesso em: 29 jun. 2010.

Contato

Eveline Batista Rodrigues
Mestre em Engenharia Elétrica – USP
Especialista em Sistemas para Internet – UNIVEM
Licenciada em Informática – FATEC
Graduada em Ciência da Computação – UNIVEM
Profª da ETEC Prof. Massuyuki Kawano – Tupã/SP.
Av. Hygino Muzzi Filho, 529 - Campus Universitário - Cep 17525-901 - Cx.Postal 2041. Fone: (14) 2105-0800 - Marília/SP
Email: evelinerodrigues@yahoo.com

José Eduardo Santarem Segundo
Doutor e Mestre em Ciência da Informação / UNESP
Especialista em Desenvolvimento de Sistemas para Internet
Docente do Centro Universitário Eurípides Soares da Rocha - UNIVEM
Coordenador do Núcleo de Prática em Sistemas de Informação - COMPSI/UNIVEM
Prof. Associado Faculdade de Tecnologia de Lins - FATEC/Lins
Av. Hygino Muzzi Filho, 529 - Campus Universitário - Cep 17525-901 - Cx.Postal 2041. Fone: (14) 2105-0800 - Marília/SP
E-mail: santarem@univem.edu.br