

Estruturação de dados semânticos com base em um modelo de repositório digital

Lucas Carmichael dos Santos de Oliveira

Faculdade de Tecnologia de Garça - carmichael.lucas@outlook.com

Larissa Pavarini da Luz

Universidade Estadual Paulista Júlio de Mesquita Filho - Marília

Faculdade de Tecnologia de Garça – larissa.luz01@fatec.sp.gov.br

Resumo

Aplicações digitais são desenvolvidas com objetivo para atender uma necessidade específica, gerando fins lucrativos ou não, essas aplicações apenas pessoas são capazes de entender o significado dos dados, através de um raciocínio e uma busca própria. Muitos destes projetos são desenvolvidos sem fornecer semântica dos dados, ou até mesmo trabalhar com conceitos de metadados, onde os agentes computacionais possam interagir e entender os significados dos dados ali existente. O objetivo deste trabalho é mostrar a importância da aplicação semântica e estruturação dos dados, para que esses agentes ao buscar um determinado termo ou informação em nossa aplicação possa entender o significado dos dados trafegados. Para o desenvolvimento deste trabalho foi realizado levantamento bibliográfico de natureza exploratória, para fins de atingir a proposta de aplicação semântica e de estruturação dos dados no projeto, onde além da interação de pessoas, as máquinas passam a entender o significado dos dados e interagir com mais clareza em no repositório digital proposto. Ao final do trabalho é apresentado o desenvolvimento de uma aplicação estruturada de forma semântica através do uso do enriquecimento dos dados em um documento HTML utilizando assim, conceito de estruturação dos dados com Microdata.

Palavras-chave: Microdata. Repositório Digital. Semântica. Estruturação de dados.

Structuring semantic data based on a digital repository model

Abstract

Digital applications are developed with the objective of meeting a specific need, generating profitable or not, these applications only people are able to understand the meaning of the data, through reasoning and an own search. Many of these projects are developed without providing data semantics, or even working with metadata concepts, where computational agents can interact and understand the meanings of the data there. The objective of this work is to show the importance of semantic application and data structuring, so that these agents, when searching for a certain term or information in our application, can understand the meaning of the data being transferred. For the development of this work, an exploratory bibliographic survey was carried out, in order to achieve the proposal of semantic application and data structuring in the project, where in addition to the interaction of people, the machines come to understand the meaning of the data and interact with more clarity in the proposed digital repository. At the end of the work, it is presented the development of a semantically structured application through the use of data enrichment in an HTML document using the concept of data structuring with Microdata..

Keywords: Microdata. Digital Repository. Semantics. Structuring of data.

1 Introdução

A Web em seu início, foi definida para troca de informações entre pessoas, que possuíam computadores para colaborar com a comunicação, auxiliando assim em buscas e trabalhos diversos.

Segundo Vidotti (2016, p.44),

[...] Podemos pensar na Internet como uma grande biblioteca, ou como um ambiente hipermídia coletivo, no qual os usuários são agentes ativos do processo de armazenamento, indexação, recuperação e disseminação de documentos eletrônicos hipertextuais, um ambiente auto-organizado em permanente mutação.

Segundo a afirmação, Vidotti (2001, p.44) afirma que computadores oferecem no seu contexto atual, o auxílio de forma limitada ao armazenamento, processamento e recuperação da informação, deixando funções de extração e interpretação dos dados a seres humanos.

A Web Semântica hoje visualizada pela composição de dados e por isso chamada de Web de dados, consiste em um conjunto de padrões destinados a fazer com que o material publicado na Web possa ser recuperado de forma Semântica, agrupando informações com o mesmo significado, independentemente de sua estrutura sintática, de forma a permitir a associação de termos, que são relacionados a estrutura cerebral do ser humano, porém de difícil relacionamento nos sistemas de informação (SANTAREM, 2010, p.16).

Para melhor similaridade da Semântica, o exemplo da sentença apresentada a seguir:

- Rex é um dinossauro;
- Existiu um dinossauro de nome Rex.

Visualiza-se que cada uma das sentenças descritas, pode ser classificada como um conjunto de dados, que unidos formam uma determinada informação. Porém se refletir sobre ambas as expressões, informam o mesmo significado. Pode-se afirmar então, que ambas são semanticamente equivalentes, apesar de serem sintaticamente diferentes.

As pessoas possuem a capacidade de entender seu significado, que “Rex” é um nome dado a um “dinossauro”, sendo ele um animal já extinto, porém o processo de recriar esta habilidade que as pessoas possuem de inferir conhecimentos através de sentenças em agentes computacionais é o principal objetivo da Web Semântica, ou seja, permitir a criação de padrões para que todo o conteúdo da Web possa ser entendido não somente por humanos, mas também por agentes computacionais (BERNES-LEE *et al.*, 2001).

Ao pensar em estruturação dos dados, cria-se uma analogia com repositórios para organização de informações. Assim nesse contexto, repositórios digitais são coleções de informações, podendo ser construídos de diferentes formas e com diferentes propósitos. A **Figura 1** retrata um exemplo real de um repositório.

Figura 1 – Exemplo da página do Wikipédia.



Fonte: Wikipédia – A enciclopédia livre (2020)

Os Repositórios Digitais (RDs), segundo o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT, 2018),

[...] são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática. Os RD's armazenam arquivos de diversos formatos e ainda, resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionando uma maior visibilidade aos resultados de pesquisas e possibilitando a preservação da memória científica de sua instituição.

A criação de um repositório digital classifica os conceitos de estruturação dos dados e conceitos de uso vinculados a Web Semântica. Utilizado na abordagem deste artigo o ambiente institucional das Faculdades de Tecnologias (Fatecs) é o foco o estudo, porém em o trabalho trabalha com dados da FATEC da cidade de Garça para demonstração dos resultados.

2 Justificativa

O estudo apresentado neste artigo inicialmente enfatiza a representação dos dados estruturados com conceitos aplicados para Web Semântica em um modelo de Repositório Digital. Todavia, com o desenvolvimento deste Repositório Digital pode validar e consequentemente nos deparar com uma maior compreensão do que é a Web Semântica e o quão a Ciência da Computação corrobora e relaciona-se com contribuições em conjunto à Ciência da Informação para que tais Ciências possam caminhar em passos semelhantes na busca por disseminar conhecimento e pesquisas relacionadas ao futuro da Web.

3 Objetivo

A pesquisa relatada neste artigo tem como objetivo geral propor um modelo de repositório digital que utilize os conceitos e as tecnologias da Web Semântica para aprimorar o enriquecimento dos dados e transpassar significado das informações à agentes computacionais.

A partir desse objetivo, visou-se estender o processo de estruturação dos dados com a utilização da Web Semântica.

Os objetivos específicos delimitados no desenvolvimento do trabalho:

- Definição das relações existentes entre as propriedades da Microdata;
- Relacionar as propriedades do Microdata com os elementos HTML;
- Implementação de uma ferramenta que valide o modelo proposto.

4 Desenvolvimento

A ferramenta proposta por este artigo foi desenvolvida com o pensamento de um projeto escalável utilizando tecnologias *open source*. A escolha das ferramentas e metodologias para elaboração do projeto são mais bem detalhadas nas próximas seções.

4.1 Python

A linguagem de programação Python foi escolhida para implementar a estrutura base do projeto, possuindo uma sintaxe clara e concisa que favoreceu a legibilidade do código-fonte, tornando o projeto simples e escalável.

Python é uma linguagem de programação interpretada, orientada a objetos, de alto nível e com semântica dinâmica. Criada no início dos anos 90, por Guido van Rossum a linguagem mostra simplicidade e reduz a manutenção do projeto. Suportada pela maioria dos sistemas operacionais e possui a capacidade de auxiliar outras linguagens. Programas como Dropbox, Reddit e Instagram são exemplos de projetos escritos em Python (Python.org, 2020).

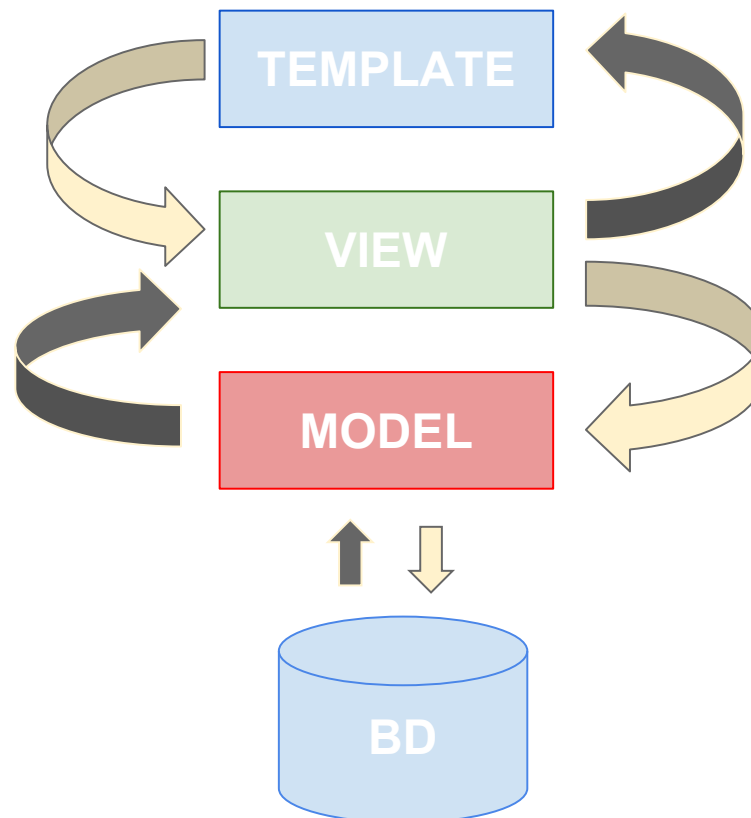
4.2 Django Framework

Django é um *framework* para aplicações Web, escrito em Python. Um *framework* é um conjunto de componentes que ajuda a desenvolver aplicações de forma mais rápida, fácil e organizada, segundo (DJANGO, 2020).

Sua estrutura é organizada possuindo similaridade com o modelo MVC (*Model, View, Controller*) sendo um padrão de projeto conhecido que é o *design pattern* (padrões de projetos) baseado em separação em três camadas os quais estão conectadas entre si. A arquitetura MTV (*Model, Template, View*) é uma derivação deste padrão de projeto, e o que se altera é a nomenclatura de arquivos e quais as camadas estão interconectadas, e este modelo é utilizado dentro do *framework* Django.

A definição de *Model* é a mesma encontrada na arquitetura MVC, que é o arquivo que contém a estrutura lógica do projeto e funciona como uma ponte para manipular dados entre o banco de dados e as *Views*. Dentro do arquivo *Model* é determinado quais tipos de dados, e como será armazenado dentro do banco de dados proposto no trabalho, como será exibido e quando for requisitado pelas *Views*. A Figura 2, apresenta a ideia de funcionamento do Framework Django utilizando a estrutura MVT.

Figura 2 – Estrutura MVT do Django Framework.



Fonte: Própria do autor (2020).

A camada *View* é muito confundida com o *Controller* da arquitetura MVC, porém eles não tem nenhuma relação. O papel desta camada é formatar os dados que estão transitando do banco de dados através da *Model* para visualização.

Já a camada *Template* é responsável pela visualização do usuário final, e se transforma como se fosse o *front-end* da aplicação. A camada arquitetura armazena os arquivos HTML, CSS, JavaScript, e tem como funcionalidade auxiliar a velocidade de maior desempenho e conforme o retorno da aplicação, renderizando os arquivos HTML da aplicação no navegador.

4.3 HTML

HTML (*HyperText Markup Language*) é a principal linguagem de marcação da *World Wide Web* projetado como uma linguagem para descrever semanticamente documentos na Web, como por exemplo, trabalhos científicos. Seu *design* geral, no entanto, permite que seja adaptado para descrever vários outros tipos de documentos e até aplicativos (W3C, 2020).

A definição e estruturação de um documento HTML visa conceitos semânticos, porém não é implementado em sua maioria, ao abrir *sites* e visualizar seus códigos-fonte verifica uma estrutura muito distinta da Figura 3, onde é exemplificado um padrão a ser seguido no desenvolvimento de qualquer aplicação.

Figura 3 – Estrutura Semântica HTML.

```
<body>
  <header>
    <nav>
      <!--NAVEGAÇÃO SITE-->
    </nav>
  </header>

  <main>
    <section>
      <article>
        <h1>Representação de Títulos</h1>
        <p>representação de paragrafos</p>
      </article>
    </section>

    <aside>
      <article>
        <h1>Representação de Títulos</h1>
        <p>representação de paragrafos</p>
      </article>
    </aside>
  </main>

  <footer>
    <article>
      <h1>Representação de Títulos</h1>
      <p>representação de paragrafos</p>
    </article>
  </footer>
</body>
```

Fonte: Própria do autor (2020).

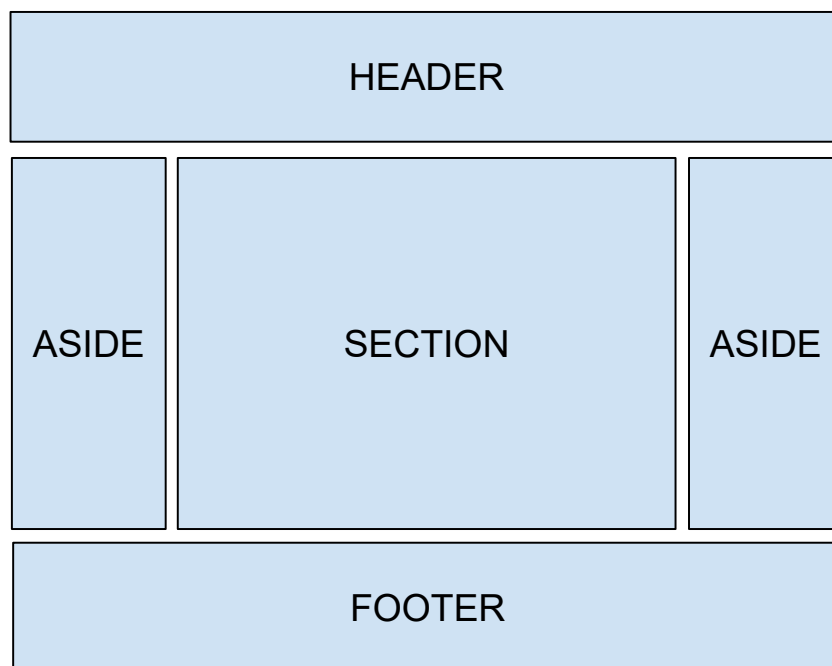
A representatividade dos elementos apresentado na Figura 3, é importante para garantir o mínimo de semântica nas aplicações desenvolvidas para Web, e os pontos a seguir é explicado de forma breve o conceito de cada elemento representado:

- *Body*, o mesmo que traduzido do inglês como “corpo”, é onde permanece o conteúdo que se disponibiliza para usuário final, um exemplo é, permitir a entrada em qualquer aplicação, o que permite a visualização e é o que fica incluso no interior desta *tag*;
- *Header*, é onde se encontra o cabeçalho do corpo da página da aplicação desenvolvida, semanticamente ele é responsável por ter *menus*, logotipo do *site* e barras de navegação;
- *Nav*, denominado semanticamente como um navegador da aplicação é responsável por delimitar área de navegação;
- *Main*, é um *tag* semântica para organizar o código-fonte desenvolvido, sua função determina o principal conteúdo dentro do *body*;
- *Section*, responsável por definir uma sessão de conteúdo na página do *site*.
- *Aside*, representa uma seção de uma página que consiste no conteúdo que é tangencialmente relacionado ao conteúdo do seu entorno, que pode ser considerado separado do conteúdo;
- *Article*, representa uma composição independente em um documento, página, aplicação, ou *site*, ou que é destinado a ser distribuído de forma independente ou reutilizável;

- *Footer* representado como rodapé da aplicação, responsável por acompanhar *links* úteis de pesquisa, ou até mesmo como a tag se comporta no navegador para indicar barras de navegação.

Contudo, existe outros elementos HTML com representatividade semântica, que fazem com que máquinas entendam o significado da informação daquela região mapeada, porém ainda não consegue definir o que nela a informação representa. A Figura 4 exibe um exemplo de uma página conceitualmente representada para entendimento humano.

Figura 4 – Estrutura Semântica representado em um site.



Fonte: Própria do autor (2020).

4.4 Microdata

O *Microdata* é um padrão de representação de informações que estende as potencialidades semânticas do HTML e assim garante a Estruturação dos Dados. As informações representadas através de *Microdata* utilizam elementos pertencentes a um vocabulário específico que deve ser referenciado no documento HTML (FONTE).

O vocabulário é a especificação das propriedades utilizadas e qual tipo de informação elas representam.

A especificação de *Microdata* define um mecanismo que permite que agentes computacionais, por meios de acesso a documentos HTML, reconheça dados a ser incorporar dados multimídias de uma forma compreensível. Estes dados devem ser legíveis tanto para máquinas como humanos, sendo também compatíveis com outros formatos de dados existentes como RDF () ou JSON (). (FONTE)

Os *Microdata* são atributos colocados nos elementos do HTML com a responsabilidade para enriquecimento semântico dos dados. Veja a sentença, “Meu nome é Lucas”, que foi representado através da Figura 5 e sua aplicação juntamente aos elementos HTML.

Figura 5 – Aplicação de Microdata em um documento HTML.

```
<article itemscope>
  <p>
    Meu nome é
    <span itemprop="name">Lucas</span>
  </p>
</article>
```

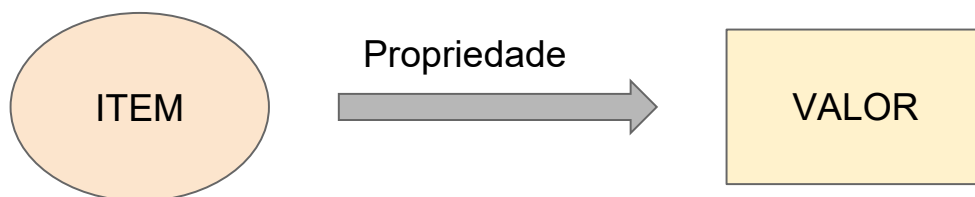
Fonte: Própria do autor (2020).

Por sua vez, quando este documento estiver disponível na Web, agentes computacionais serão capazes de identificar através de buscas, e entender que na sentença citada, existe uma pessoa cujo, nome é “Lucas”.

Por exemplo, os mecanismos de pesquisa podem identificar melhor o conteúdo da página usando as anotações do *Schema.org*, que será o conjunto de vocabulários que foi utilizado no desenvolvimento da aplicação, e os sistemas de gerenciamento de conteúdo podem encontrar e usar as informações dos documentos, uma vez que estiverem marcadas de maneira conhecida por esta base.

O *Microdata* fornece um mecanismo simples para rotular o conteúdo de um documento, para que possa ser processado como um conjunto de itens descritos por par de item-valor, como apresentado na Figura 6.

Figura 6 – Representação de um par nome-valor.



Fonte: Própria do autor (2020).

5 Resultados

Na busca de comprovar a ferramenta proposta nesta pesquisa, foi implementado uma aplicação capaz de executar os principais pontos discutidos no artigo, utilizando os métodos e ferramental apresentado.

Metodologicamente, tal aplicação tem a função de ser uma prova de conceito, em que as principais contribuições da pesquisa são expostas e demonstradas. Como relatado, essa prova de conceito foi construída com o intuito de validar o modelo, visando demonstrar a aplicação sendo compreendida por agentes computacionais. A ferramenta ainda não tem a função de ser uma implementação final da proposta, pois essa não é a meta desta pesquisa, mas sim ser um reagent na validação dos objetivos propostos.

A Figura 7 é demonstrado partes do projeto desenvolvido e validado por meio de uma ferramenta de teste de dados estruturado disponibilizado pelo Google, a fim de comprovar aplicação Semântica e Estruturação dos Dados.

Neste contexto ainda é apresentado um teste que verifica a aplicação Semântica e Estruturação dos dados, onde é possível ver que agentes computacionais através do enriquecimento do código HTML do projeto, foi aplido em todas as partes do projeto e

Estruturação de dados semânticos com base em um modelo de repositório digital

clarificamos o significado das informações trafegadas para que possa ser compreendido por máquinas e humanos.

Figura 7 – Ferramenta de teste de dados estruturados disponibilizado pelo Google.

The screenshot displays the Google Structured Data Testing Tool. On the left, the HTML code is shown with annotations for semantic elements. On the right, a table lists the extracted structured data:

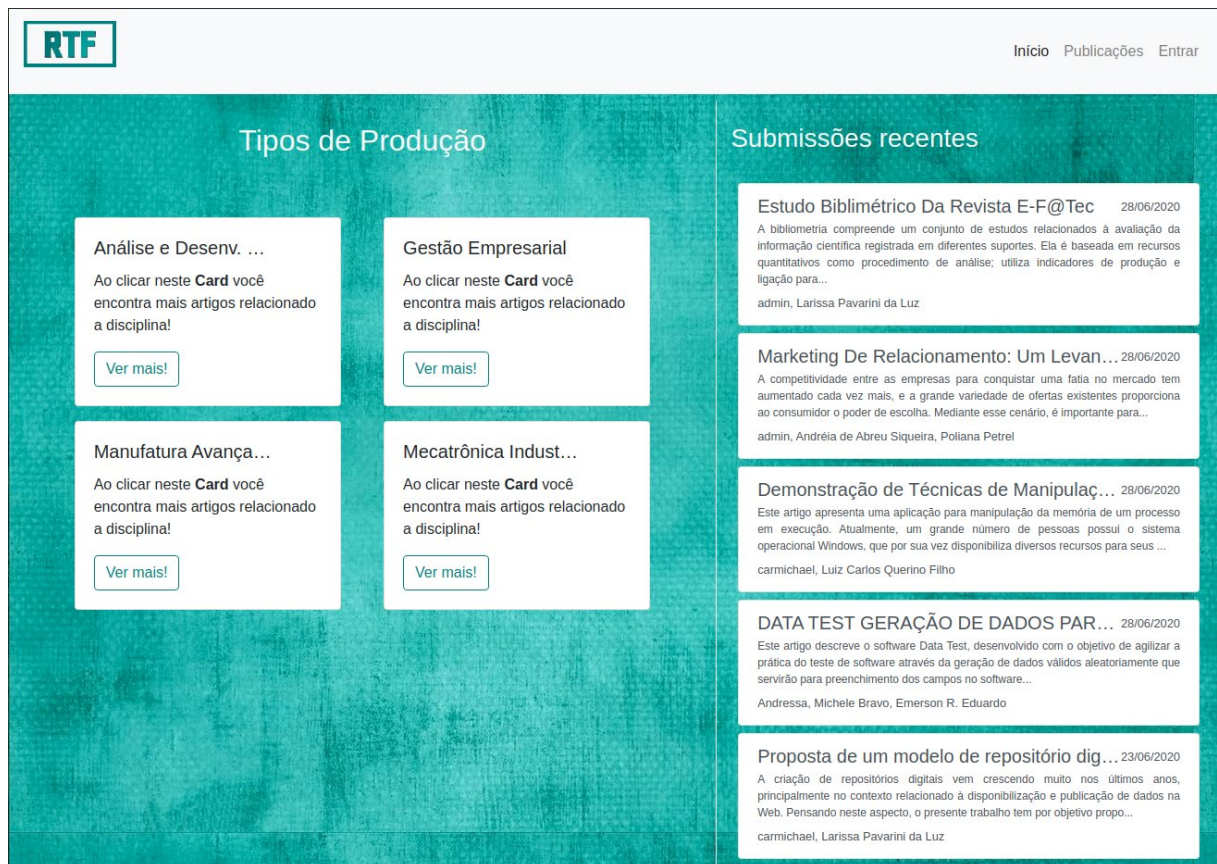
@type	ScholarlyArticle
image	https://search.google.com/static/img/logo.png
headline	Proposta de um modelo de repositório digital utilizando conceitos da Web Semântica
abstract	A criação de repositórios digitais vem crescendo muito nos últimos anos, principalmente no contexto relacionado à disponibilização e publicação de dados na Web. Pensando neste aspecto, o presente trabalho tem por objetivo propor a criação de um repositório digital para a publicação de dados científicos com conceitos de Web Semântica. Este repositório terá como objetivo ampliar e disseminar as atividades de pesquisas desempenhadas por pesquisadores ou terceiros afins de gerar, divulgar e preservar a sua produção científica. Neste sentido, o trabalho apresenta uma revisão bibliográfica de natureza exploratória, partindo de conceitos iniciais, e uma proposta para construção de um Repositório Digital a partir da coleta de registros das diversas fontes que encontramos disponíveis na Internet, com o intuito de facilitar o uso dos dados publicados. Para isso, propõe-se a aplicação dos conceitos, fundamentos e boas práticas da Web Semântica e do Linked Data para permitir o enriquecimento, armazenamento e recuperação de informações de forma semântica e ligada, a partir de um vocabulário aberto.
datePublished	2020-06-23
genre	Artigo
sameAs	https://search.google.com/media/documents/2020/06/document.pdf
about	
@type	Thing
name	Publicação!
author	
@type	Thing
name	carmichael
author	
@type	Thing
name	Larissa Pavarini da Luz
maintainer	
@type	Thing

Fonte: Própria do autor (2020).

Entretanto, o modelo do repositório apresentado consisti três tipos de telas principais: 1) “Início”, em que o usuário verifica tipos de produção disponíveis e submissões recentes. 2) “Publicações”, em que o usuário verifica as publicações referente a um item de produção escolhido. 3) “Artigo”, em que exhibe para o usuário a publicação escolhida para melhor apreço do trabalho publicado.

A primeira tela, demonstrada na Figura 8, contém os tipos de produção separados por cursos referentes a instituição, neste exemplo, a Fatec Garça. Em que o usuário pode buscar publicações por itens de coleções ou verificar as submissões recentes. A inserção da estruturação dos dados e aplicação da semântica ocorre na exibição dos dados e por toda a página.

Figura 8 – Tela inicial do Repositório Digital.



Fonte: Própria do autor (2020).

A Figura 9 representa as relações existentes de publicações na base de dados criada, referente a coleção selecionada pelo usuário, na tela chamada de “Publicações”, em que é sinalizado as publicações em si e o campo de busca para alguma obra específica, o usuário seleciona a publicação que deseja. Os termos apresentados representam as relações que foram localizadas entre o termo de busca escrito pelo usuário com aplicação semântica dos dados e estrutura referente a utilização dos vocabulários mantidos pela Schema.org.

Figura 9 – Tela de publicações de um item de produção.



The screenshot shows a web interface for a digital repository. At the top, there is a logo with the letters 'RTF' in a teal box. Below the logo, the title 'Publicações referente ao item de coleção: Análise e Desenv. de Sistemas' is displayed. A search bar with the placeholder text 'Buscar pelo Título ?' and a 'Buscar' button is located below the title. The main content area contains four publication entries, each with a title, a date, a short abstract, and the author's name.

RTF

Publicações referente ao item de coleção: Análise e Desenv. de Sistemas

Buscar pelo Título ?

Estudo Bibliométrico Da Revista E-F@Tec 28/06/2020
A bibliometria compreende um conjunto de estudos relacionados à avaliação da informação científica registrada em diferentes suportes. Ela é baseada em recursos quantitativos como procedimento de análise; utiliza indicadores de produção e ligação para...
, Larissa Pavarini da Luz

Demonstração de Técnicas de Manipulação de Memóri... 28/06/2020
Este artigo apresenta uma aplicação para manipulação da memória de um processo em execução. Atualmente, um grande número de pessoas possui o sistema operacional Windows, que por sua vez disponibiliza diversos recursos para seus ...
, Luiz Carlos Querino Filho

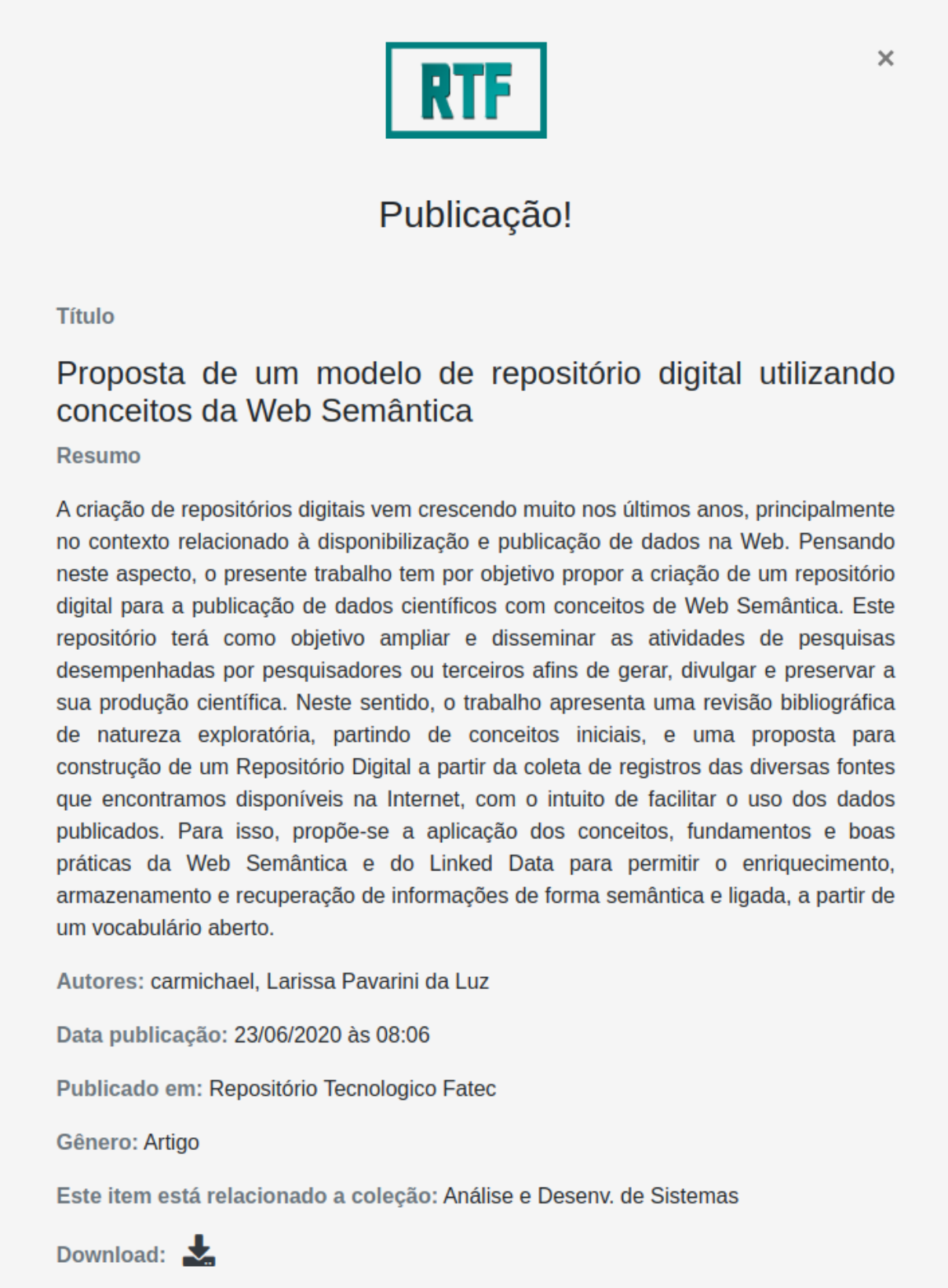
DATA TEST GERAÇÃO DE DADOS PARA O APOIO D... 28/06/2020
Este artigo descreve o software Data Test, desenvolvido com o objetivo de agilizar a prática do teste de software através da geração de dados válidos aleatoriamente que servirão para preenchimento dos campos no software...
, Michele Bravo, Emerson R. Eduardo

Proposta de um modelo de repositório digital utilizando... 23/06/2020
A criação de repositórios digitais vem crescendo muito nos últimos anos, principalmente no contexto relacionado à disponibilização e publicação de dados na Web. Pensando neste aspecto, o presente trabalho tem por objetivo propo...
, Larissa Pavarini da Luz

Fonte: Própria do autor (2020).

Por fim, na Figura 10, tem-se o objetivo de visualizar o artigo escolhido pelo usuário com o objetivo de identificar o trabalho de pesquisa ou relacionado.

Figura 10 – Tela de exibição do artigo buscado pelo usuário.



RTF X

Publicação!

Título

Proposta de um modelo de repositório digital utilizando conceitos da Web Semântica

Resumo

A criação de repositórios digitais vem crescendo muito nos últimos anos, principalmente no contexto relacionado à disponibilização e publicação de dados na Web. Pensando neste aspecto, o presente trabalho tem por objetivo propor a criação de um repositório digital para a publicação de dados científicos com conceitos de Web Semântica. Este repositório terá como objetivo ampliar e disseminar as atividades de pesquisas desempenhadas por pesquisadores ou terceiros afins de gerar, divulgar e preservar a sua produção científica. Neste sentido, o trabalho apresenta uma revisão bibliográfica de natureza exploratória, partindo de conceitos iniciais, e uma proposta para construção de um Repositório Digital a partir da coleta de registros das diversas fontes que encontramos disponíveis na Internet, com o intuito de facilitar o uso dos dados publicados. Para isso, propõe-se a aplicação dos conceitos, fundamentos e boas práticas da Web Semântica e do Linked Data para permitir o enriquecimento, armazenamento e recuperação de informações de forma semântica e ligada, a partir de um vocabulário aberto.


Autores: carmichael, Larissa Pavarini da Luz

Data publicação: 23/06/2020 às 08:06

Publicado em: Repositório Tecnológico Fatec

Gênero: Artigo

Este item está relacionado a coleção: Análise e Desenv. de Sistemas

Download: 

Fonte: Própria do autor (2020).

Contudo, afirma-se que o uso de conceitos da Web Semântica e Estruturação em todas as páginas apresentadas, dando significado às informações desde, conteúdos acadêmicos, como as publicações, mas também a aplicação da estruturação em menus de navegação, busca e informações de rodapé.

6 Conclusão

A natureza interdisciplinar da computação permite que diversos estudos abarcados por essas áreas estejam vinculadas por outras disciplinas na busca de solucionar questões relativas à informação. Nesse âmbito, a estruturação dos dados e aplicação semântica, mostra a relevância e importância do enriquecimento semântico. Fornece um meio para aprimorar a forma como motores de buscas e outros sistemas que utilizam da recuperação de dados para consumo, satisfaz e entrega mais valor ao usuário final.

A evolução dos estudos tratando da Web Semântica está criando um arcabouço teórico acerca dos conceitos e das tecnologias dessa proposta, que tem como consequência a execução de estudos teóricos e pragmáticos capazes de trazer contribuições fundamentais em diversos campos de estudos, principalmente no que tange à compreensão do contexto de um determinado domínio. Essa intersecção entre a estruturação dos dados e a Web Semântica tem como objetivo fornecer propriedade para os mecanismos computacionais compreenderem aspectos relativos a uma determinada busca, que por vezes eram compreendidos somente por humanos.

Vale destacar ainda que a implementação desta proposta demonstrou a viabilidade dessa forma de interação entre o sistema e os usuários, e sistema e outras máquinas.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v.284, n.5, p.34-43, 2001.

CRISTOVÃO, H. M.; FERNANDES, J. H. C. Recuperação de informação em dados ligados: um modelo baseado em mapas conceituais e análise de redes complexas. *Transinformação*, v. 30, n. 2, p. 193-207, 2018. Disponível: <http://dx.doi.org/10.1590/2318-08892018000200005>. Acesso em: 15 set. 2019.

DJANGO. Documentação. Disponível eletronicamente em <<https://docs.djangoproject.com/pt-br/3.1/faq/models/>> . Acesso em 09 dez. 2020.

EVANS, P.; WURSTER, T. Getting Real About Virtual Commerce. *Harvard Business Review*, November-December 1999.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). Repositórios Digitais. Disponível: <http://www.ibict.br/informacao-para-a-pesquisa/repositorios-digitais#repositorios-brasileiro>. Acesso em: 10 set 2019.

HEERY, R.; ANDERSON, S. Digital Repositories Review. Joint Information Committee. University of Bath. <http://www.ukoln.ac.uk/repositories/publications/review-200502/digital-repositories-review-2005.pdf>. Acesso em: 22 Set. 2019.

LINKED DATA. Linked Data - Connect Distributed Data across the Web. Disponível em: <http://linkeddata.org/>. Acesso em: 11 set 2019.

MARGARYAN, A.; CURRIER, S.; LITTLEJOHN, A.; NICOL, D. Learning communities and repositories. Disponível em: <https://www.gcu.ac.uk/cd-lor//learningcommunitiesreport.pdf>. Acesso em: 22 set 2019.

NHACUONGUE, J. A.; ROZSA, V.; LIMA DUTRA, M. Linked Data e Ciência da Informação: diretrizes para a publicação de datasets institucionais abertos. *Biblios* [online]. 2018, n.73, pp.20-34. ISSN 1562-4730. Disponível em: <http://dx.doi.org/10.5195/biblios.2018.429>.

OPENREFINE. A free, open source, powerful tool for working with messy data. Disponível em <http://openrefine.org/>. Acesso em: 22 Set. 2019.

PREECE, Jennifer; ROGERS, Yvonne; SHAP, Helen. Design de Interação: além da interação homem-computador. 1. ed. Porto Alegre : Bookman, 2005.

PRESSMAN, Roger S. Software e Engenharia de Software in Engenharia de Software. São Paulo: Pearson Makron Books, 2005.

PYTHON. Python 3.9.1 documentation. Disponível em <https://docs.python.org/3/>. Acesso em: 22 Set. 2019.

RODRIGUES, R., S.; TAGA, V.; VIERIA, E. M. F.; Repositórios Educacionais: estudo preliminares para a Universidade Aberta do Brasil. Disponível em: http://www.brapci.inf.br/_repositorio/2015/01/pdf_d7c05dcbb5_0027522.pdf. Acesso em 22 Set. 2019.

ROZSA, V.; DUTRA, M.; NHACUONGUE, J. Linked Open Data no contexto acadêmico: identificação e análise de vocabulários utilizados na academia e na pesquisa científica. *Brazilian Journal of Information Science: research trends*. v. 11, n. 3, 9 out. 2017.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, v. 21, n. 2, p. 90-115, 2016a. Disponível em <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/27939/20122>. Acesso em: 14 set. 2019.

SANTAREM SEGUNDO, J. E. Representação Iterativa: um modelo para Repositório Digital. 2010. “bibliografia 140-150 f.” Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010.

SENSO, J. A.; ARROYO MACHADO, W. (2018). La publicación en Linked Data de registros bibliográficos: modelo e implementación. *Revista Española de Documentación Científica*, 41 (4): e217. Disponível em: <https://doi.org/10.3989/redc.2018.4.1535>.

SOMMERVILLE, Ian. Engenharia de Software. São Paulo: Pearson Prentice Hall, 2011

VIDOTTI, S. A. B. G. et al. Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT. In: Encontro Nacional de Pesquisa em Ciência da Informação, 17., 2016, Bahia, 2016. Anais... Bahia: ANCIB; UFBA, 2016, p. 1-21. Disponível em: <http://hdl.handle.net/11449/144718>.

VIDOTTI, S. A. B. G. et al. Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. Disponível em:

<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/388/932>.
Acesso em: 22 Set. 2019

WIKIPÉDIA – A enciclopédia livre (2020). Disponível eletronicamente: <
https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:P%C3%A1gina_principal>. Data de acesso out.
2020.