

# Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Lightbase  
Consultoria em Software Público

eduardo.santos@lightbase.com.br  
eduardo.edusantos@gmail.com

www.eduardosan.com  
www.lightbase.com.br

19 de Outubro de 2012

# Para começar

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Para entender a Web Semântica é preciso **esquecer** os conceitos de publicação de conteúdo na Internet.
- Web Semântica envolve **necessariamente** entender uma nova tecnologia.
- Conceitos da Web Semântica [W3C, 2012]
  - 1 Dados Linkados
  - 2 Vocabulários e Ontologias
  - 3 Consultas
  - 4 Inferência
  - 5 Aplicações Verticais

## Sumário

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso
- 5 Como armazenar?
- 6 Algumas aplicações
- 7 Referências

## 1 Introdução à Web Semântica

- Web 2.0
- Web Semântica

## 2 O formato de dados da Web Semântica

## 3 Armazenando os dados

## 4 Caso de uso

## 5 Como armazenar?

## 6 Algumas aplicações

## 7 Referências

# Um pouco da história da Internet



A visão Google das coisas

# O que foi Web 2.0

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
**Web 2.0**  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

*Web 2.0 é uma série de aplicações que propiciam e potencializam a formação de redes sociais digitais*

Abel Reis

# Ferramentas da Web 2.0

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações



# E aí?

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- A quantidade de conteúdo gerado pelas redes é simplesmente **IMENSA!**
- Os dados no ciberespaço [Santos, 2012] estão infinitamente dispersos.
- Exemplo:
  - Onde está sua agenda de trabalho?
  - Onde estão seus amigos?
  - Onde estão seus contatos de trabalho?
  - Onde está você no ciberespaço?



# O mundo dos dados

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- A Internet das coisas
- O espalhamento do eu
- A consolidação das informações públicas

# Como funciona?

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Os dados precisam ser catalogados de uma maneira **diferente do habitual**
- Exemplo trazido da língua portuguesa: estrutura de triplas

Sujeito Quem?

Predicado O quê?

Objeto Quanto? Quando? (Valores)

# Um exemplo conceitual

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica

Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

*O Brasil ganhou três medalhas de ouro em Londres  
2012*

**Sujeito** Brasil Londres 2012

**Predicado** Medalhas de ouro

**Objeto** 3 medalhas

# Uma outra abordagem

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

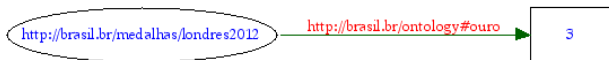
Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

*O Brasil ganhou três medalhas de ouro em Londres 2012*



Uma outra visão sobre a mesma informação no formato de um grafo

# O dado dentro do dado

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Princípio básico: o dado deve ser **auto-contido** ou **auto-descrito**.
- Descrever o dado faz com que ele esteja acessível tanto por **homens** quanto por **máquinas**.
- Parte fundamental da descrição do dado: **Vocabulários** e **Ontologias**.

# O dado dentro do dado

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Princípio básico: o dado deve ser **auto-contido** ou **auto-descrito**.
- Descrever o dado faz com que ele esteja acessível tanto por **homens** quanto por **máquinas**.
- Parte fundamental da descrição do dado: **Vocabulários** e **Ontologias**.
- Guarde essa ideia: o dado deve ser **auto-contido**

# Vocabulários e Ontologias

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Como saber que o dado se refere a medalhas de ouro?
- Como saber o que é uma olimpíada ou que se trata de Londres 2012?
- O dado é gramaticalmente correto? Ex.: Faz sentido dizer que medalhas de ouro em Londres foram feias?
- É possível descrever até mesmo uma quantidade finita de valores para o objeto (tipos de dado).

# Descrevendo Vocabulários e Ontologias

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Como saber quem é o sujeito?
  - Acessando o endereço  
<http://www.brasil.br/medalhas/londres2012>
- Como saber do que estamos falando?
  - Acessando o endereço  
<http://brasil.br/ontology#ouro>
- O objeto expressa o dado propriamente dito, e não precisa de descrição.
  - Quantidade: **3**



- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso
- 5 Como armazenar?
- 6 Algumas aplicações
- 7 Referências

# Formalizando a estrutura de triplas

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

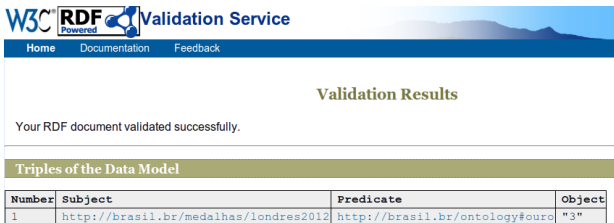
O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações



The screenshot shows the W3C RDF Validation Service interface. At the top, there is a navigation bar with links for Home, Documentation, and Feedback. The main heading is "Validation Results". Below this, a message states "Your RDF document validated successfully." A section titled "Triples of the Data Model" contains a table with the following data:

Number	Subject	Predicate	Object
1	<a href="http://brasil.br/medalhas/londres2012">http://brasil.br/medalhas/londres2012</a>	<a href="http://brasil.br/ontology#ouro">http://brasil.br/ontology#ouro</a>	"3"

Ferramenta fundamental:  
`http://www.w3.org/RDF/Validator/`

# O formato RDF+XML

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

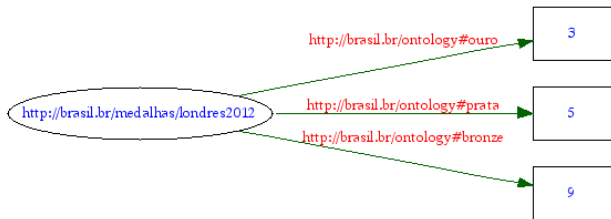
Algumas aplicações

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://brasil.br/ontology#" >
  <rdf:Description rdf:about="http://brasil.br/medalhas/londres2012" >
    <p:ouro>3</p:ouro>
  </rdf:Description>
</rdf:RDF>
```

# Incluindo as outras medalhas

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://brasil.br/ontology#" >

  <rdf:Description rdf:about="http://brasil.br/medalhas/londres2012" >
    <p:ouro>3</p:ouro>
    <p:prata>5</p:prata>
    <p:bronze>9</p:bronze>
  </rdf:Description>
</rdf:RDF>
```



# Ontologias e Vocabulários

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Ainda é preciso entender o que significa cada informação
- Como saber o que é ouro, prata e bronze?
- É possível validar os dados?

# Ontologias e Vocabulários

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://brasil.br/ontology#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://brasil.br/ontology">

  <owl:Ontology rdf:about="#">
    <rdfs:comment xml:lang="pt-br">
      <![CDATA[<p>Ontologia para descrever medalhas olimpicas</p>]]>
    </rdfs:comment>
  </owl:Ontology>

  <!-- Campos -->
  <rdf:Property rdf:ID="ouro" rdfs:comment="Medalhas de ouro">
    <rdfs:domain rdf:resource="#ouro"/>
    <rdfs:range rdf:resource="http://www.w3.org/TR/xmlschema-2/#integer">
  </rdf:Property>

  <rdf:Property rdf:ID="prata" rdfs:comment="Medalhas de prata">
    <rdfs:domain rdf:resource="#prata"/>
    <rdfs:range rdf:resource="http://www.w3.org/TR/xmlschema-2/#integer">
  </rdf:Property>

  <rdf:Property rdf:ID="bronze" rdfs:comment="Medalhas de bronze">
    <rdfs:domain rdf:resource="#bronze"/>
    <rdfs:range rdf:resource="http://www.w3.org/TR/xmlschema-2/#integer">
  </rdf:Property>

</rdf:RDF>
```

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados**
- 4 Caso de uso
- 5 Como armazenar?
- 6 Algumas aplicações
- 7 Referências

# Quebrando paradigmas

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Já podemos ver que a estrutura de triplas é **diferente** do modelo **orientado a objeto**
- Um objeto possui atributos; um sujeito possui predicados e objetos
- A definição de objeto na Web Semântica é **DIFERENTE** da definição de objeto em programação
- O modelo de dados da Web Semântica envolve registros **auto-contidos**
- Boa notícia: o modelo OO está mais próximo da Web 3.0 que o modelo relacional



# Armazenando os dados

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Premissas da Web Semântica:
  - Os dados devem ser **auto-contidos**, ou seja, todas as informações sobre um registro precisam estar juntas.
  - Os registros precisam ser **agnósticos**, ou seja, um registros pode ter uma estrutura diferente do outro.
  - Dados e metadados
- Premissas de bancos de dados relacionais:
  - Relacionamentos
  - Tabelas com linhas e colunas
  - Estrutura de dados **pré-definidas**

# Armazenando os dados

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Premissas da Web Semântica:
  - Os dados devem ser **auto-contidos**, ou seja, todas as informações sobre um registro precisam estar juntas.
  - Os registros precisam ser **agnósticos**, ou seja, um registros pode ter uma estrutura diferente do outro.
  - Dados e metadados
- Premissas de bancos de dados relacionais:
  - Relacionamentos
  - Tabelas com linhas e colunas
  - Estrutura de dados **pré-definidas**

**POLÊMICA:** O modelo relacional não é adequado para a Web Semântica!?

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso**
- 5 Como armazenar?
- 6 Algumas aplicações
- 7 Referências

# Exemplo: medalhas na olimpíada de Londres

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica

Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Objetivo: descrever a quantidade de medalhas de ouro conquistadas pelo Brasil na olimpíada de Londres.
- Formato da Web Semântica: estrutura de triplas.
  - Modelo relacional: tabelas e colunas.
  - Modelo **documental**: registros auto-contidos.

# Modelo documental

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web

Semântica

Web 2.0

Web

Semântica

O formato de dados da Web

Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://brasil.br/ontology#" >
  <rdf:Description rdf:about="http://brasil.br/medalhas/londres2012" >
    <p:ouro>3</p:ouro>
    <p:prata>5</p:prata>
    <p:bronze>9</p:bronze>
  </rdf:Description>
</rdf:RDF>
```

# Conclusões

## Banco de dados para a Web Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica

Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Para **alguns cenários** o modelo relacional é desnecessariamente complicado
- Dividir em tabelas e colunas é uma escolha do usuário?  
**Pense nisso!**
- **Pense no usuário**

# Modelando para a web semântica

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- **Importante:** os registros devem ter um identificador único.
- **URI**
- Formato básico de armazenamento: **CHAVE x VALOR**

**Chave** Identificador único do registro. Pode ser uma URI, um número, um nome, etc. Mas precisa ser único.

**Valor** Dados e metadados do registro no formato da Web Semântica

# Lightbase [Lightbase, 2012b]

- Um pouco de história:
  - Empresa de software das mais antigas do Brasil: mais de 25 anos desenvolvendo **bancos de dados**.
  - Banco amplamente utilizado em Europa e América.
  - Vários prêmios em revistas e congressos especializados
  - Um pequeno problema:

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações



# Lightbase [Lightbase, 2012b]

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Um pouco de história:
  - Empresa de software das mais antigas do Brasil: mais de 25 anos desenvolvendo **bancos de dados**.
  - Banco amplamente utilizado em Europa e América.
  - Vários prêmios em revistas e congressos especializados
  - Um pequeno problema: nossos concorrentes sempre foram grandinhos.
  - Um caso prático dos problemas da concorrência desleal:  
<http://www.eduardosan.com/2012/05/30/a-verdadeira-defesa-do-software-livre/>

# Lightbase [Lightbase, 2012b]

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Um pouco de história:
  - Empresa de software das mais antigas do Brasil: mais de 25 anos desenvolvendo **bancos de dados**.
  - Banco amplamente utilizado em Europa e América.
  - Vários prêmios em revistas e congressos especializados
  - Um pequeno problema: nossos concorrentes sempre foram grandinhos.
  - Um caso prático dos problemas da concorrência desleal:  
<http://www.eduardosan.com/2012/05/30/a-verdadeira-defesa-do-software-livre/>
- A empresa sempre desenvolveu com foco no usuário
- Deve ser possível a um usuário não técnico descrever sua própria base de dados.

# Lightbase Neo: o banco de dados para a Web Semântica [Lightbase, 2012a]

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Principais conceitos:

**Bases** Descrição da estrutura de dados através de ontologia

**Registros** Formato **LBDF** inspirado no ODF

**Formulários** Descrição das entradas de dados

**Relatórios** Formatação da apresentação dos dados

# Lightbase Neo: o banco de dados para a Web Semântica [Lightbase, 2012a]

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Principais conceitos:

**Bases** Descrição da estrutura de dados através de ontologia

**Registros** Formato **LBDF** inspirado no ODF

**Formulários** Descrição das entradas de dados

**Relatórios** Formatação da apresentação dos dados

- Para a Web Semântica o banco de dados é um conceito que pode ser implementado em **quase todos** os bancos de dados.

# Descritor de bases

- A base é descrita pelo usuário em **qualquer formato**.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE Base SYSTEM "Base.dtd">
<base>
  <nome>Pessoa</nome>
  <descricao><![CDATA[Como descrever seus amigos]]></descricao>
  <senha><![CDATA[0$!@#%fhbhfhdh54745754]]></senha>
  <objeto>
    <campo>
      <nome>nome</nome>
      <descricao><![CDATA[Esse e o nome da pessoa]]></descricao>
      <tipo>Texto</tipo>
      <indexacao>
        <indice>Palavra</indice>
      </indexacao>
    </campo>
    <campo>
      <nome>cpf</nome>
      <descricao><![CDATA[Esse e o CPF da pessoa]]></descricao>
      <tipo>Inteiro</tipo>
      <indexacao>
        <indice>Unico</indice>
      </indexacao>
    </campo>
  </objeto>
</base>
```

Exemplo: base no formato XML (sem semântica)

# Registro

## Banco de dados para a Web Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

## Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Conceito: dados e metadados
- Dados: arquivo binário, documento, imagem, etc.
- Metadado: RDF+XML

# Registro serializado

```
<?xml version="1.0" encoding="iso-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lb="http://rdf.lightbase.cc/ontology/"
  <lb:registro rdf:about="http://rdf.lightbase.cc/notas/NOTA/chave
    <lb:base rdf:resource="http://rdf.lightbase.cc/notas/NOTA/
      <lb:arquivo rdf:resource="3462012_3C3B_arquivo.doc"/>
      <lb:baseName>notas</lb:baseName>
      <dc:title xml:lang="pt-br">Notas</dc:title>
      <dc:description>Base de Notas</dc:description>
      <dc:identifier><![CDATA[3462012]]></dc:identifier>
      <lb:fields rdf:parseType="Collection">
        <lb:field rdf:about="http://rdf.lightbase.cc/not
          <lb:fieldName><![CDATA[orgao]]></lb:field
            <dc:title xml:lang="pt-br"><![CDATA[Orga
              <dc:description xml:lang="pt-br"><![CDAT
                <lb:fieldType><![CDATA[ALFANUMERIC]]></
                  <lb:fieldIndex><![CDATA[PORCAMPO | TEXTU
                    </lb:field>
                      (...)
                    </lb:fields>
                  </lb:registro>
                </rdf:RDF>
```

Exemplo: registro no formato LBDF

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso
- 5 Como armazenar?**
- 6 Algumas aplicações
- 7 Referências



# Armazenamento do registro

- Armazenamos o registro no formato CHAVE x VALOR
- Dado: imagem, documento, etc. Sistema de arquivos ou banco!
- Metadado: XML, JSON, RDF+XML, turtle, etc.

ID	Metadado	Dado
1	<pre>&lt;?xml version="1.0"?&gt; &lt;rdf:RDF&gt;   &lt;lb:registro rdf:about="teste"&gt;     (...)   &lt;/lb:registro&gt; &lt;/rdf:RDF&gt;</pre>	/srv/arquivos/teste.odt
2	<pre>&lt;?xml version="1.0"?&gt; &lt;rdf:RDF&gt;   &lt;lb:registro rdf:about="teste2"&gt;     (...)   &lt;/lb:registro&gt; &lt;/rdf:RDF&gt;</pre>	/srv/arquivos/teste2.odt

# Indexação

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- A indexação envolve **recuperar** o documento através dos metadados.
- Ex.: numa base de despachos jurídicos, ordenar os registros pelo nome do advogado

Nome do advogado	ID do registro
João	2
José	1

Importante: o índice **não é o registro**

# Busca e recuperação textual

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web

Semântica

Web 2.0

Web Semântica

O formato de dados da Web

Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Quero recuperar todos os advogados que começam com J

String de busca	ID do registro
J	2
J	1

# O elefante flexível

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- O elefante pode resolver todos os itens

**Armazenamento** Tabela com três campos: ID, dado e metadado

**Indexação** XML com xpath e índices

**Recuperação textual** Tsearch2

# Solução ideal

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

Chave x Valor PostgreSQL

Indexação Elastic Search

Busca textual Elastic Search

- O Lightbase é uma **especificação** de troca de dados no formato da Web Semântica.
- A especificação **independe** das escolhas tecnológicas descritas acima.

# Criando uma tabela com chave valor

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

```
CREATE TABLE registros (  
    chave serial PRIMARY KEY,  
    metadado xml NOT NULL,  
    dado varchar NOT NULL  
);  
  
CREATE INDEX id_registro  
ON registros USING BTREE  
( CAST(xpath('/lb:registro/text()', metadado) AS TEXT[]) );
```

# Cria índice no Elastic Search

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web

Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

```
curl -XPUT 'http://ckan.brighthouse.net/api/data/pareceres/_settings' -d '{
  "index.analysis.analyzer.default.type":"custom",
  "index.analysis.analyzer.default.tokenizer":"standard",
  "index.analysis.analyzer.default.filter.0":"lowercase",
  "index.analysis.analyzer.default.filter.1":"asciifolding",
  "index.analysis.analyzer.default.filter.2":"pt_stemmer",
  "index.analysis.filter.pt_stemmer.type":"stemmer",
  "index.analysis.filter.pt_stemmer.name":"portuguese"
}'

curl -s -XPUT 'http://ckan.brighthouse.net/api/data/pareceres/_mapping' -d '{
  "pareceres" : {
    "properties" : {
      "data" : {
        "type": "date",
        "index": "not_analyzed"
      }
    }
  }
}
```

Exemplo de indexação no Elastic Search

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso
- 5 Como armazenar?
- 6 Algumas aplicações**
- 7 Referências



# Aplicação

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Depois de tudo isso, faltou o principal: pra que serve tudo isso?
- Riqueza: extrair informações pertinentes sobre os dados

# Carolina Dieckmann

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

## Carolina Dieckmann usa cílios enormes para clipe das Olimpíadas

A atriz deu vida a Afrodite em 'Os Deuses do Olimpo'.

do EGO no Rio

12 comentários

Tweetar 42

Recomendar 28

### SAIBA MAIS

Veja as fotos dos bastidores da gravação do clipe "Os Deuses do Olimpo"

acontecerá no Rio de Janeiro.

A gravação, dirigida pelo marido da atriz **Regina Casé**, Estevão Ciavatta, da Pindorama Filmes, aconteceu na quadra da Portela, em Madureira, subúrbio do Rio. A música que embala o videoclipe foi composta por Arlindo Cruz, Rogê e Arlindo Neto e é cantada por Diogo Nogueira, Mart'nália, Mr. Catra, Thalma de Freitas, **Zeca Pagodinho**, Ed Motta e o próprio Arlindo nos vocais e muitos outros.

Além de Carolina, integram também a filmagem . **Fernanda Monteneegro**. Martinho da Vila.

Para viver Afrodite, no clipe "Os Deuses do Olimpo", Carolina Dieckmann ganhou enormes cílios postiços. O clipe, exibido a partir desta segunda-feira, 13, marca a contagem regressiva para as Olimpíadas 2016 que

### Seus amigos no EGO

veja o que eles estão lendo



Conecte-se com Facebook

Conecte-se com seus amigos e saiba o que eles estão lendo. [Veja mais](#)

### TUDO SOBRE



**Carolina Dieckmann**

[página da famosa »](#)



**Veja o making of de 'Os Deuses do Olimpo' visitam o Rio de Ja...**

há 1 hora



**Confira as imagens que marcaram a semana de 4 a 10 de agosto**

há 2 dias

- Diferentes aplicações em diferentes ecossistemas podem trocar informações
- SPARQL: um canal de comunicação **INDEPENDENTE** de tecnologia
- No contexto da Web Semântica, é necessário um banco de dados?
- Dados Abertos

# Aplicações com dados abertos

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

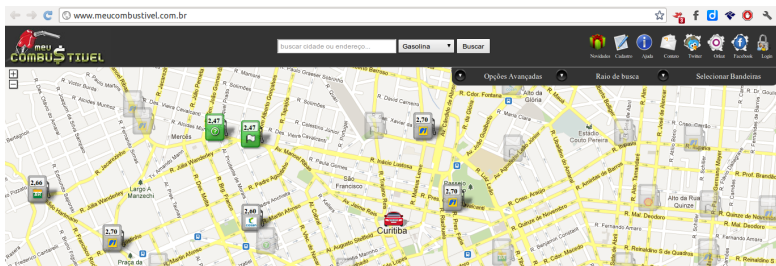
O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações



[www.meucombustivel.com.br](http://www.meucombustivel.com.br)

# Trabalhando com dados abertos

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

Nome	Qty.	Média
BANCO CENTRAL DO BRASIL	4459	R\$ 16.737,60
ADVOCACIA-GERAL DA UNIAO	7546	R\$ 15.088,26
CENTRO NAC.TECNO.ELETRONICA AVANCADA S.A	7	R\$ 14.374,32
CONTROLADORIA-GERAL DA UNIAO	2342	R\$ 14.247,22
INSTITUTO DE PESQUISA ECONOMICA APLICADA	564	R\$ 14.184,95
SUPERINTENDENCIA DE SEGUROS PRIVADOS	422	R\$ 13.678,75
MINISTERIO DA FAZENDA	32596	R\$ 12.487,48
EMPRESA DE PESQUISA ENERGETICA	8	R\$ 12.402,34
AGENCIA NACIONAL DE AGUAS	328	R\$ 12.379,99
COMISSAO DE VALORES MOBILIARIOS	564	R\$ 12.352,41
EMPRESA DE TRENS URBANOS DE PORTO ALEGRE	13	R\$ 11.327,64
AGENCIA NAC PETROLEO GAS NAT BIOCOMBUSTI	685	R\$ 11.301,15
MINISTERIO DAS RELACOES EXTERIORES	1547	R\$ 11.129,46
AGENCIA NACIONAL DE VIGILANCIA SANITARIA	1934	R\$ 10.952,20
AGENCIA NACIONAL DE SAUDE SUPLEMENTAR	602	R\$ 10.804,12

<http://blog.girino.org/2012/07/15/fucando-a-transparencia-parte-2/>

# Oportunidades

Banco de  
dados para a  
Web  
Semântica

Eduardo  
Ferreira dos  
Santos

Introdução à  
Web  
Semântica  
Web 2.0  
Web  
Semântica

O formato  
de dados da  
Web  
Semântica

Armazenando  
os dados

Caso de uso

Como  
armazenar?

Algumas  
aplicações

- Libere os dados e descubra o que as pessoas vão fazer com eles!
- Descubra o que fazer com os dados e ganhe dinheiro com isso.
- Reflexão: e o que estão fazendo com seus dados?

# LBW Publica

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

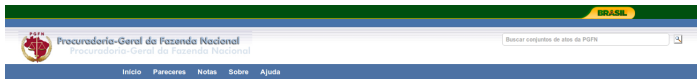
O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações



Conjuntos de atos da PGFN

[Pareceres](#)

[Notas](#)

`dados.pgfn.gov.br`

- O serviço exporta dados do Lightbase para o formato RDF
- Um *daemon* recolhe os dados e importa no Ckan
- Uma nova interface indexa o conteúdo dos pareceres através do Elastic Search

# Indexação e recuperação textual

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica Web 2.0 Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

Buscar Pareceres da PGFN

Q id: 4572011

1 documentos encontrados

## Parecer N° 457/2011

data	28/03/2011
orgao	PGFN/CAT
procurador	FRANCISCO TARGINO DA ROCHA NETO
ementa	Dúvidas suscitadas pela Superintendência da Receita Federal do Brasil em Minas Gerais sobre os termos do Acordo celebrado entre a União, o Estado de Minas Gerais e o INSS nos autos do Recurso Especial nº 1.135.162, a respeito de contribuições para o RGPS relativas a servidores não efetivos da referida unidade da federação.
Baixar arquivo	<a href="http://dados.pgfn.fazenda.gov.br:80/storage/2012-07-09T110449/Parecer_0457-2011.doc">http://dados.pgfn.fazenda.gov.br:80/storage/2012-07-09T110449/Parecer_0457-2011.doc</a>

## PARECER PGFN/CAT Nº 457/2011

Dúvidas suscitadas pela Superintendência da Receita Federal do Brasil em Minas Gerais sobre os termos do Acordo celebrado entre a União, o Estado de Minas Gerais e o INSS nos autos do Recurso Especial nº 1.135.162, a respeito de contribuições para o RGPS relativas a servidores não efetivos da referida unidade da federação. Art. 48, § 13 da Constituição Federal de 1988, introduzido pela EC nº 26, de 1998. Resposta a algumas das indagações.

dados.pgfn.gov.br



# Lightbase Neo

Banco de dados para a Web Semântica

Eduardo Ferreira dos Santos

Introdução à Web Semântica  
Web 2.0  
Web Semântica

O formato de dados da Web Semântica

Armazenando os dados

Caso de uso

Como armazenar?

Algumas aplicações

- Nova tecnologia para armazenamento e distribuição de informações no formato da Web Semântica
- Bases, formulários e relatórios
- Portabilidade e inspiração no ODF: LBDF
- Ainda em processo de desenvolvimento
- Acompanhe nosso trabalho: <http://rdf.lightbase.cc>
- Conheça o Lightbase: <http://www.lightbase.com.br>
- Participe da comunidade:  
[http://www.softwarepublico.gov.br/ver-comunidade?community\\_id=3673574](http://www.softwarepublico.gov.br/ver-comunidade?community_id=3673574)

- 1 Introdução à Web Semântica
  - Web 2.0
  - Web Semântica
- 2 O formato de dados da Web Semântica
- 3 Armazenando os dados
- 4 Caso de uso
- 5 Como armazenar?
- 6 Algumas aplicações
- 7 Referências



Lightbase (2012a).

Lightbase e a web semântica.

<http://rdf.lightbase.cc/> Acessado em 19/10/2012.



Lightbase (2012b).

O lightbase.

<http://www.lightbase.com.br/> Acessado em 19/10/2012.



Santos, E. (2012).

Um pouco sobre a definição de ciberespaço.

<http://www.eduardosan.com/tag/ciberespaco> Acessado em 21/08/2012.



W3C (2012).

Material do w3c sobre a web semântica.

<http://www.w3c.br/Padroes/WebSemantica> Acessado em 21/08/2012.

Eduardo Ferreira dos Santos  
Lightbase Consultoria

eduardo.santos@lightbase.com.br  
eduardo.edusantos@gmail.com

www.lightbase.com.br  
www.eduardosan.com

+55 61 3347-1949