

XML

***eXtensible
Markup
Language***

Rubens Queiroz de Almeida

UNICAMP

queiroz@unicamp.br

*It is not the strongest species that
survive, nor the most intelligent,
but the one most responsive to
change*

Charles Darwin

Comentários

Uma nova linguagem de coificação de dados poderia dar aod EDI o impulso necessário para atingir milhões de usuários. A combinação de EDI com XML oferece a promessa de estender as vantagens de EDI baseado na Web, através de uma especificação padrão, a milhões de pequenas e médias empresas

Iniciativas de comércio baseadas em documentos como as do grupo EDI/XML, se bem sucedidas, darão a desenvolvedores de aplicações e integradores novas ferramentas para criar um novo enfoque a EDI que não é possível hoje.

Comentários (...)

O padrão XML provavelmente será uma parte importante do futuro do comércio eletrônico

Por muitos anos o EDI tem prometido revolucionar a forma com que as empresas se comunicam e trocam dados umas com as outras. Com XML/EDI, esta promessa poderá finalmente se tornar uma realidade

Comentários (...)

XML torna possível uma infinidade de formatos para intercâmbio de dados. XML é o estilo da Web para EDI

XML pode fazer pelos dados o que a linguagem Java fez pelos programas, que é tornar os dados independentes de plataforma e fornecedores de software

... uma linguagem obscura chamada XML está a ponto de revolucionar o ciberespaço

Comentários (...)

As implicações mais amplas de um formato de dados padrão e de fácil processamento como XML, são difíceis de imaginar, mas certamente irão representar um grande impacto no comércio eletrônico. É claro hoje que o comércio eletrônico irá em breve se tornar sinônimo de comércio em geral

A Internet Hoje

- Primordialmente um meio de acesso a imagens e textos
- Não existem padrões para buscas inteligentes, intercâmbio de dados, apresentação de informação de modo adaptável e personalização da informação
- Embora tenha todo tipo de informação a Internet assemelha-se a uma enorme biblioteca sem catalogação com todos os livros espalhados no chão

A Internet Hoje (...)

- Um tremendo potencial para negócios
 - CISCO
 - Amazon
- Estrutura inadequada para seu crescente uso comercial

O Culpado

- O grande culpado por esta situação é justamente o padrão HTML
- Este padrão, hoje amplamente difundido, não é capaz de mudar esta situação pois é um formato que descreve como uma página Web deve ser exibida e não oferece nenhuma descrição dos dados

Documentos HTML

```
<HTML>
```

```
<HEAD>
```

```
<meta name="Author" content="Rômulo Queiroz Almeida">
```

```
</HEAD>
```

```
<BODY>
```

```
<H1>Título 1 </H1>
```

```
<H2>Título 2</H2>
```

```
<H3>Título 3</H3>
```

```
<P>Texto Normal</P>
```

```
<UL style="list-style-type: none;">
```

```
<LI>Item 1</LI>
```

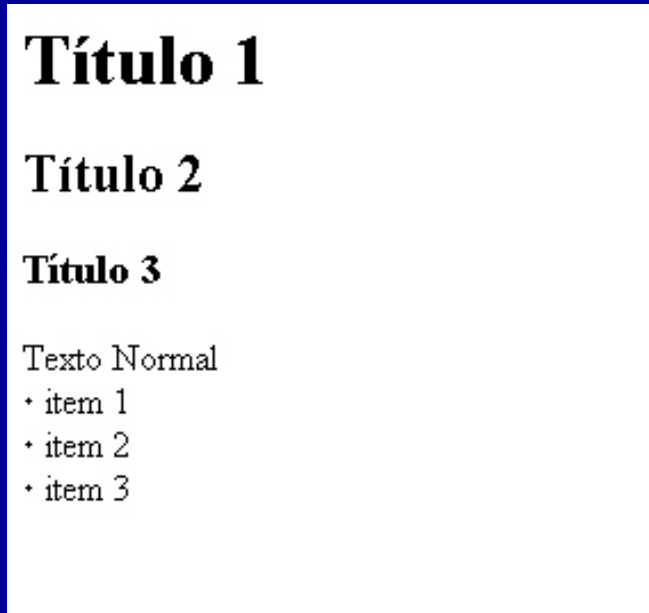
```
<LI>Item 2</LI>
```

```
<LI>Item 3</LI>
```

```
</UL>
```

```
</BODY>
```

```
</HTML>
```



Razões da Popularidade do Padrão HTML

- Extremamente simples
- Estilo próprio para formatação de documentos
- Ligações de hipertexto criadas facilmente
- Suporte a formulários
- Interação homem x máquina
- Programação simples
 - CGI (Common Gateway Interface)

Limitações do Padrão HTML

- Estrutura limitada e simples
 - número fixo de tags
 - projetado para formatação de documentos entregues através da Web
- Difícil reutilização da informação
- Padrão limitado para intercâmbio de informações

Limitações do Padrão HTML

(...)

- Inadequado para organização das informações. Busca por documentos gera grande número de resultados irrelevantes
- Automação limitada
- Padrão modificado constantemente
- Adequação aos novos padrões requer um grande esforço

Standard Generalized Markup Language *(SGML)*

- Padrão internacional, definido em 1986, para formato de texto e documentos
- Popular em organizações que precisam criar e gerenciar grandes volumes de documentos
- Padrão adotado, entre outros, pela indústria aero-espacial, automotiva, de telecomunicações e de software
- Compatível com padrões atuais e futuros
- Não proprietário, não se torna obsoleto
- Formato texto, liberta o conteúdo de produtos e softwares

SGML e HTML

- HTML é uma aplicação derivada do padrão SGML
- HTML oferece um conjunto limitado de tags
- SGML oferece a possibilidade da criação de qualquer conjunto de tags

SGML: Definição de Padrões

- Os representantes de um setor se associam e criam um padrão comum para intercâmbio de dados
- O padrão comum (DTD - Document Type Definition) cobre o conjunto de tags a serem adotadas e a gramática de uma linguagem de marcação (markup language)

Padrões Adotados

- JCALS - Departamento de Defesa
- J-2008 - Setor Automotivo e de transporte
- ATA-100 - Indústria aeroespacial
- EDB - Indústria de Semicondutores
- DocBook - Indústria de informática

SGML - Problemas

- Difícil adoção na Web
 - Complexidade
 - torna difícil seu suporte por browsers web
 - Inexistência de estilos amplamente difundidos
 - Para uso na web é convertido para HTML perdendo muito da inteligência do documento original, impedindo a sua reutilização, intercâmbio e automação

- SGML para a Web
- Formato para descrição de dados estruturados
- Otimizado para entrega de informações SGML através da Web

XML (...)

- Dados estruturados utilizando XML são independentes dos aplicativos utilizados ou fornecedores de software
- Suporta praticamente todas as funcionalidades mais difundidas do padrão SGML
- Simplex:
 - especificação SGML: 500 páginas
 - Especificação XML: 33 páginas

Documentos XML

- Formato texto, semelhante ao HTML em muitos aspectos
- Permite um número ilimitado de tags, cada uma delas indicativa, não de como algo deve ser exibido, e sim do que significa

Criadores XML

- Padrão em desenvolvimento por um grupo de trabalho do W3C (WWW Consortium)
- Este grupo é constituído por cerca de 14 empresas e organizações, entre elas Adobe, Microsoft, HP, Netscape.

Apoio ao Padrão XML

- Internet Explorer e Netscape Navigator já oferecem suporte ao formato XML
- Microsoft Office 2000
- Assunto preferido da imprensa especializada
- Apenas o apoio das empresas integrantes do grupo de trabalho representa um grande passo para sua adoção universal

Perspectivas

- Tecnologia preferencial para intranets e extranets
- Torna possível a adoção da Internet como meio preferido para aplicações de negócios

Limitações SGML na Web

- Falta de suporte pelos browsers mais populares
- Uso SGML na Web envolve a tradução para HTML
 - perda de informações
 - intercâmbio e automação de dados ficam muito mais difíceis
 - impossível reconstruir o arquivo original a partir do arquivo HTML
 - Exemplo: conversão de um arquivo no formato CAD para GIF
- Falta de suporte para folhas de estilo

XML = SGML adaptado à Web

- O padrão XML foi criado para tornar possível a entrega de informações SGML na Web
- Elimina as limitações do padrão SGML ao mesmo tempo em oferece todos os seus benefícios

XML e HTML

XML

- Define o conteúdo (dados)
- Tags descrevem os dados, como temperatura, humidade, etc.
- Tags definidas pelo criador do documento
- Apresentação definida por folhas de estilo
- Dados separados da apresentação e do processamento dos dados

HTML

- Descreve o formato de apresentação
- Número limitado e não extensível de tags
- Inadequado para gerenciamento de grande volume de dados
- Não oferece a funcionalidade requerida pelo comércio eletrônico

XML: Padrões Associados

- XSL - eXtensible Style Language
 - Compatível com CSS (Cascading Style Sheet)
 - Capacidade de reordenação da informação sem consultar o servidor Web
 - Suporte aos formatos impressos e online
- XLL - eXtensible Linking Language
 - Compatível com endereçamento baseado em URL's
 - Links bi-direcionais
 - Links indiretos
 - Fim às mensagens *404 File Not Found*

Criação de Aplicações XML

- Para criar aplicações XML são necessários 4 passos:
 - Selecionar ou escrever um DTD
 - Criar documentos XML
 - Interpretar documentos XML
 - Exibir documentos XML

Seleção ou Criação DTD

- Document Type Definition ou DTD define
 - as tags de cada documento
 - quais tags podem conter outras tags
 - o número e sequência das tags
 - os atributos que as tags podem ter e seus valores

A Criação de um DTD

- Requer análise de documentos
 - que elementos irão ocorrer?
 - como se relacionam entre si?
 - como irão os usuários interagir com eles?
- Tarefa difícil
 - quanto mais amplo o escopo maior a complexidade
 - difícil conseguir a concordância de todos os envolvidos

DTDs (...)

- Empresas que desejarem usar XML precisam descobrir uma maneira simples de encontrar a informação de que necessitam sobre os esquemas (DTDs), documentos e processos de negócios que outras empresas suportam
- Este problema pode ser minimizado pela criação de portais especializados em localizar, gerenciar e informar sobre o padrão XML, XLS e os modelos de informação usados em milhares de aplicações. A Microsoft já criou um portal com este objetivo (<http://www.biztalk.org>)

Geração de um documento

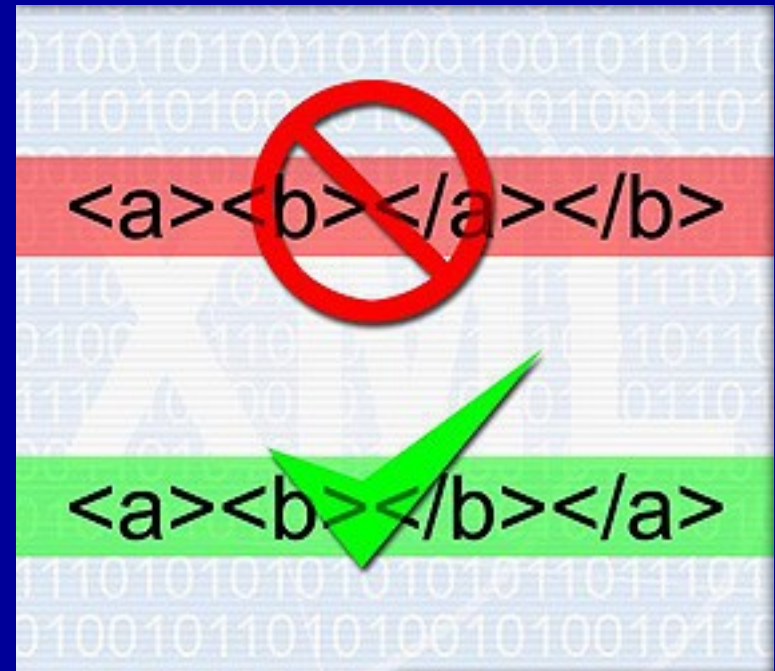
XML

Um DTD é como um formulário padrão que é preenchido. Os dados podem vir de uma consulta a um banco de dados, de uma busca em documentos ou pesquisa em um catálogo online. Quando o formulário estiver preenchido, ele pode então ser enviado a quem solicitou o documento

Regras para documentos XML

- Tags XML não podem se sobrepor
- Tags de finalização não podem ser omitidas
- Tags que não contêm nenhum texto podem conter o marcador de final na tag de início.

</BR> é equivalente a



Tipos de Documentos

- Inválidos
- Bem formados
- Válidos



Interpretação de Documentos XML

XML

- Para a interpretação de documentos XML existem duas APIs que podem ser utilizadas: DOM (Document Object Model) e SAX (Simple API for XML)
 - DOM é um padrão definido pelo W3C que cria uma visão em árvore do documento XML



Exibição de Documentos XML

Existem várias maneiras de se visualizar documentos XML. Se o browser entende o padrão XML, o documento pode ser enviado diretamente para ele. Ou então pode-se usar uma folha de estilo para transformar o arquivo XML em algo que o browser entenda

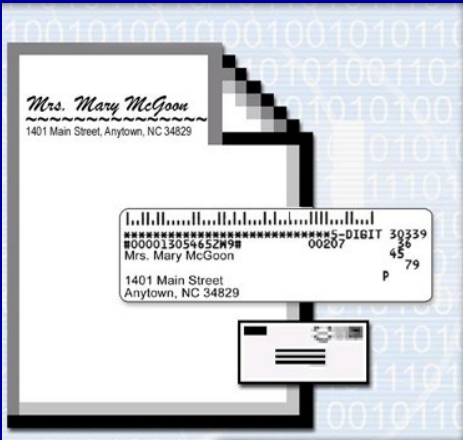


Diferentes Visões dos Mesmos Dados

```
<address>
<name>
<title>Mrs.</title>
<first-name>Mary</first-name>
<last-name>McGoon</last-name>
</name>
<street>1401 Main Street</street>
<city>Anytown</city>
<state>NC</state>
<zipcode>34829</zipcode>
...
</address>
```

```
<address>
<name>
<title>Mrs.</title>
<first-name>Mary</first-name>
<last-name>McGoon</last-name>
</name>
<street>1401 Main Street</street>
<city>Anytown</city>
<state>NC</state>
<zipcode>34829</zipcode>
...
</address>
```

Mrs. Mary McGoon
1401 Main Street
Anytown, NC 34829



Diferentes Visões dos Mesmos Dados



Aplicações XML

- O uso do XML é possível em aplicações totalmente diferentes daquelas hoje usadas por meio do padrão HTML:
 - Transações financeiras
 - Registros médicos
 - Companhias aéreas
 - Seguros ...
- XML é um formato de dados flexível capaz de representar uma grande variedade de informações

Exemplos de Aplicação

- Estruturação dos Dados (tags)
 - Fabricação de torneiras
 - Aposento (cozinha, banheiro, tanque)
 - Controles (torneira, alavanca)
 - Acabamento (cromado, bronze)
 - Preço (barato, moderado, caro)
 - Uma planilha de dados para cada torneira
- Sistema automatizado para obter os dados de um banco de dados tradicional e outras informações a partir de um banco de dados de documentos

Exemplo de Aplicação - Vantagens

- Automação
Informações mantidas em websites podem ser criadas de forma totalmente automática
- Buscas precisas
A adoção de um padrão comum irá facilitar a busca de informações por parte dos clientes



Exemplo de Aplicação - Vantagens

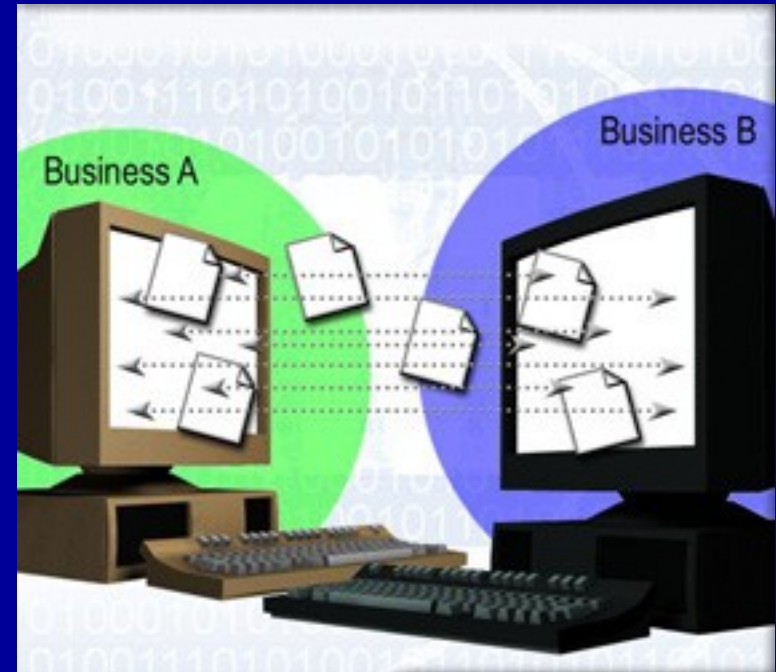
(...)



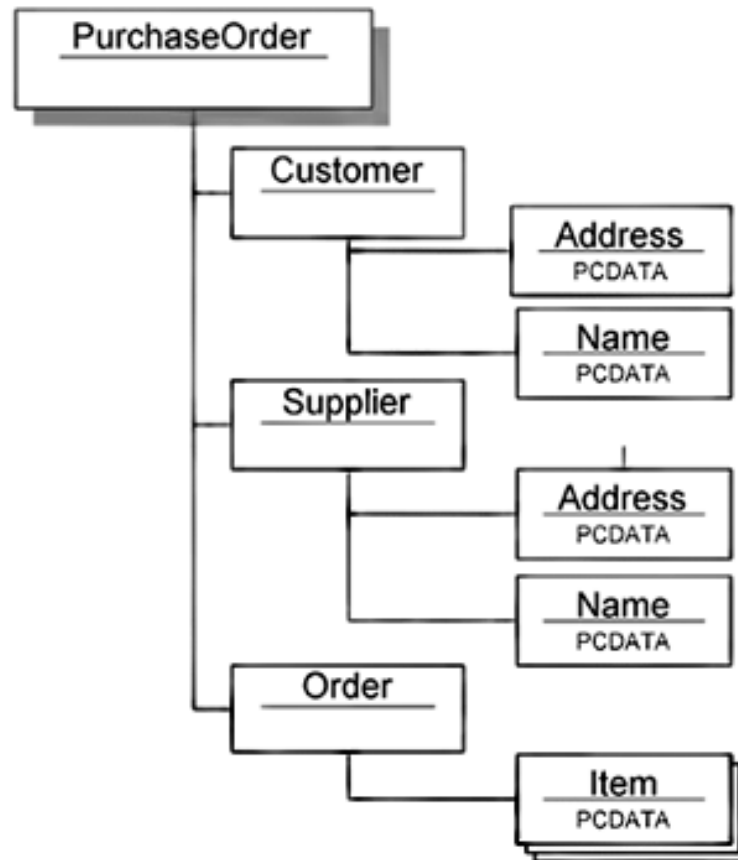
- Intercâmbio de dados
construtores podem obter informações sobre os produtos da Web e importar estes dados diretamente em seus aplicativos de cálculo de custo e projeto sem nenhum esforço adicional
- Imune a mudanças em browsers web
formatação definida em folha de estilos separada do documento

Comunicação entre Empresas

Com XML, a comunicação entre empresas torna-se muito mais simples com a adoção de um padrão comum de codificação de documentos (DTD)



Ordem de Compra



Ordem de Compra

```
<?xml version='1.0' ?>
  <PurchaseOrder>
    <Customer>
      <Name>Acme Airline and Stormdoor Company</Name>
      <Address>34 Poplar St., Watertown, MA</Address>
    </Customer>
    <Supplier>
      <Name>Computer And Printer Paradise</Name>
      <Address>45 Seabreeze Dr., Miami, FL</Address>
    </Supplier>
    <Order>
      <Item>A brown hat</Item>
    </Order>
  </PurchaseOrder>
```

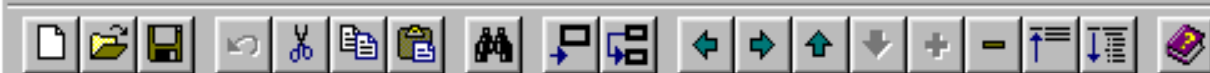


Structure	Values
PURCHASEORDER	
CUSTOMER	
NAME	Acme Airline and Stormdoor C...
ADDRESS	34 Poplar St., Watertown, MA
SUPPLIER	
NAME	Computer And Printer Paradise
ADDRESS	45 Seabreeze Dr., Miami, FL
ORDER	
ITEM	A brown hat

Compras Online

```
<?xml version="1.0"?>
<!DOCTYPE booklist SYSTEM
"book.dtd">
<Booklist>
<Book>
  <Title>Eiger Dreams</Title>
  <Author>Krakauer, Jon</Author>
  <Deal>
<Merchant>Amazon.com</Merchant>
  <Format>Paperback</Format>
  <Price>$10.36</Price>
  <Availability>1 day</Availability>
</Deal>
<Deal>
  <Merchant>A1books</Merchant>
  <Format>paperback</Format>
  <Price>$10.00</Price>
  <Availability>2-3
days</Availability>
</Deal>
```

```
<Deal>
  <Merchant>A1books</Merchant>
  <Format>audio</Format>
  <Price>$17.50</Price>
  <Availability>2-3
days</Availability>
</Deal>
<Deal>
  <Merchant>All
Direct</Merchant>
  <Format>paperback</Format>
  <Price>$9.71</Price>
  <Availability>3
days</Availability>
</Deal>
</Book>
</Booklist>
```

Structure	Values
BOOKLIST	
BOOK	
TITLE	Eiger Dreams
AUTHOR	Krakauer, Jon
DEAL	
MERCHANT	Amazon.com
FORMAT	Paperback
PRICE	\$10.36
AVAILABILITY	Ships in 1 day
DEAL	
MERCHANT	A1books
FORMAT	paperback
PRICE	\$10.00
AVAILABILITY	2-3 days
DEAL	
MERCHANT	A1books
FORMAT	audio
PRICE	\$17.50
AVAILABILITY	2-3 days
DEAL	
MERCHANT	All Direct
FORMAT	paperback
PRICE	\$9.71

Benefícios

- Acesso mais fácil à informação através de atribuição de significado mais relevante aos dados
- Desenvolvimento de aplicações Web mais flexíveis
- Integração de dados de origens diferentes
- Computação e tratamento local dos dados
- Múltiplas visões dos dados
- Atualizações granulares

XML: Ferramentas

- Microsoft XML Notepad
 - Disponível gratuitamente (por enquanto)
 - Permite visualização da estrutura dos dados

Conclusões

- Por que usar XML?
 - Viabiliza a troca de dados entre computadores
 - XML é a linguagem da Web para intercâmbio de dados
 - Grande número de aplicações
 - Preparação para o mundo conectado e Internet onipresente

Conclusões (...)

- O que fazer?
 - Aprender o máximo possível sobre XML, (ainda está em tempo)
 - Analisar as necessidades cautelosamente
 - XML possui um enfoque diferenciado para resolução de problemas em relação às tecnologias tradicionais
 - Acompanhar a evolução do padrão XML
 - Investigar XML hoje e implantá-lo amanhã

Referências

- W3C
<http://www.w3c.org>
- XML.ORG
<http://xml.org>
- XML.COM
<http://www.xml.com>
- Links para outros documentos
<http://www.Dicas-L.unicamp.br/hotlinks/XML>