



Web, HTML, XML

World-Wide Web

- A World-Wide Web (também chamada Web ou WWW) é, grosso modo, uma interface gráfica para a Internet.
- A idéia é criar um universo de informações sem fronteiras, com as seguintes características:
 - interface consistente;
 - incorporação de um vasto conjunto de tecnologias e tipos de documentos;
 - "leitura universal" (de qquer sistema).
- Para isso, implementa três ferramentas importantes:
 - um protocolo de transmissão de dados - HTTP;
 - um sistema de endereçamento próprio - URL;
 - uma linguagem de marcação, para transmitir documentos formatados através da rede - HTML.

HTTP

- HTTP (Hyper Text Transfer Protocol): É um protocolo de transferência de Hipertexto.
- Utilizado para transmissão de dados no sistema World-Wide Web.
- Cada vez que você aciona um link, seu browser realiza uma comunicação com um servidor da Web através deste protocolo
- O HTTP está sobre o protocolo de transporte TCP, que é um protocolo orientado à conexão.

HTTP

- Quando um usuário clica em um hiperlink, o navegador executa uma série de etapas em ordem para buscar a página indicada
 - O navegador determina o URL (verificando o que foi selecionado)
 - O navegador pergunta ao DNS qual é o endereço IP de `www.ufrj.br`
 - O DNS responde com `146.164.41.20`
 - O navegador estabelece uma conexão TCP com a porta 80 em `146.164.41.20`
 - Em seguida, o navegador envia um comando solicitando o arquivo `/index.html`
 - O servidor `www.ufrj.br` envia o arquivo `/index.html`
 - A conexão TCP é encerrada
 - O navegador exibe todo o texto de `/index.html`
 - O navegador busca e exibe todas as imagens que o arquivo contém

Navegação



URL

- URL - Uniform Resource Locator (Localizador Uniforme de Recursos): é baseado em uma sintaxe chamada URI - Universal Resource Identifier (Identificador Uniforme de Recursos) que serve para localizar recursos de uma maneira uniforme.
- Exemplo de URL:
 - <http://www.dcc.ufrj.br/index.html>

URL

- O endereço identifica:
 - Protocolo de acesso ao recurso desejado (http);
 - Máquina a ser conectada (www.dcc.ufrj.br);
 - Caminho do diretório (/docentes);
 - Recurso/arquivo a ser obtido (index.html).

HTML

- HTML - HyperText Markup Language (Linguagem de Marcação de Hipertexto)
- Serve para indicar formatações para textos, inserir imagens e ligações de hipertexto.
- Os navegadores são responsáveis por identificar as marcações em HTML e apresentar os documentos conforme o que foi especificado por essas marcações.

HTML

- *Hyper Text Markup Language*
 - Especificação definida pelo consórcio W3C:
<http://www.w3.org/>
- Um arquivo html contém marcadores (tags)
 - Estes marcadores indicam para o navegador (browser) como a página deve ser apresentada
 - Marcadores usualmente vem em pares: <tag>...</tag>
 - Os documentos em HTML são arquivos ASCII comuns, que podem ser editados em qualquer editor simples, apesar de existirem também editores específicos

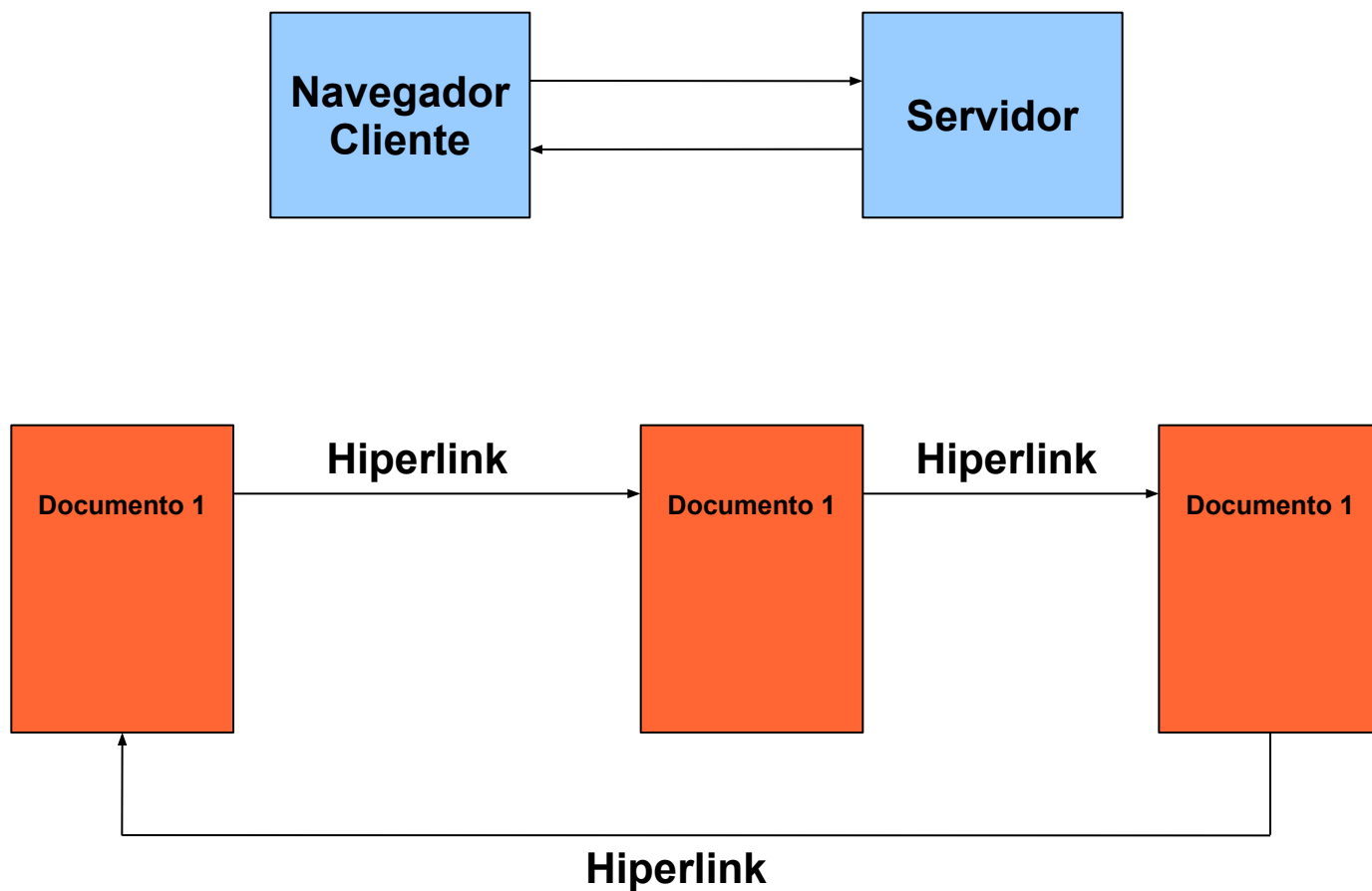
HTML

- Uma página web pode ser vista como uma coleção de elementos (palavras, parágrafos, listas, tabelas, imagens, etc.) visuais
- HTML define a estrutura e layout dos elementos em uma página Web através de tags.
- Cada tag pode ter atributos que modificam a aparência ou layout do elemento visual contido na tag.

Tags HTML

- Formato: `<Inicio formato>um texto</Fim formato>`
 - `um texto em negrito`
 - `<H1>um heading</H1>`
- Tags de Elementos vazias
 - Ex: `<HR>` `
`
- Comentários
 - `<!-- Olá, sou um comentário... -->`

Introdução



Publicação de documentos

- Para que uma página esteja permanentemente disponível pela Web, ela precisa ter um endereço fixo, alojada em um servidor.
- Definida a hospedagem, basta enviar para o provedor os arquivos de seu site (via FTP ou por uma página de envio no próprio provedor de espaço) e suas páginas já estarão disponíveis para visitas.

Problema

- Dados e apresentação misturados!
 - Como saber o que significa o que?
- XML [eXtensible Markup Language - linguagem de marcadores extensível]
 - Linguagem de marcação
 - Para *transferir*, não mostrar, *dados*
 - Tags *não* são predefinidas – autor define as tags
 - É auto-descritiva

Por que XML?

- HTML foi projetado para MOSTRAR documentos em um browser e nada mais
 - É difícil mostrar documentos em dispositivos móveis ou traduzir o conteúdo para outro idioma
- Propósito do HTML é permitir a criação rápida de documentos web. XML pode ser usado em diferentes contextos que podem não ter nada a ver com interação com humanos (ex: serviços web usam XML)
- HTML raramente dá informação sobre a estrutura do documento ou o que ele significa
 - HTML é uma linguagem de apresentação, XML é uma linguagem de descrição de dados

XML

- Não “faz” nada

- Foi criado para estruturar, armazenar e transportar informação

```
<recado>  
<para>João</para>  
<de>Maria</de>  
<assunto>Lembrete</assunto>  
<corpo>Não esqueça do chopp na sexta!</corpo>  
</recado>
```

- A mensagem é auto-descritiva: tem remetente e destinatário, assunto e corpo...
 - Informação pura é “embrulhada” em Tags

XML: eXtensible Markup Language

□ *Extensível*

- Permite que o autor defina suas próprias tags, sua ordem e como devem ser processadas ou mostradas.

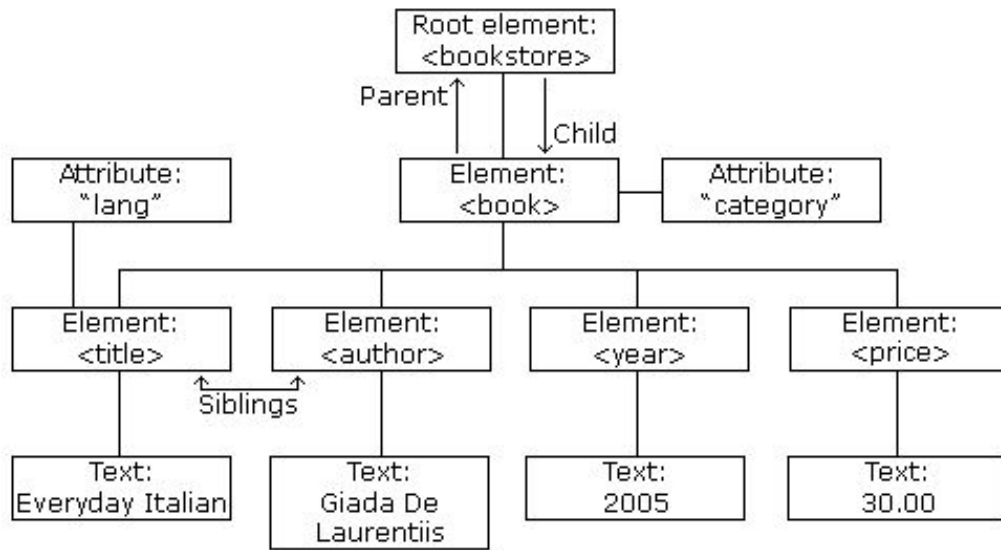
□ *Marcação*

- Característica mais reconhecível: tags ou elementos. Muito similar as tags do HTML.

□ *Linguagem*

- Linguagem similar ao HTML, porém mais flexível porque permite a criação das suas próprias tags.

Elementos formam uma árvore



```
<bookstore>
  <book category="COOKING">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="CHILDREN">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WEB">
    <title lang="en">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```

Elementos

Declaração: versão e codificação do texto

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<recado>
```

Elemento raiz

```
<para>João</para>
```

```
<de>Maria</de>
```

Elementos-filho

```
<assunto>Lembrete</assunto>
```

```
<corpo>Não esqueça do chopp de sexta!</corpo>
```

```
</recado>
```

**Final do
Elemento raiz**

Elemento XML

- Um elemento é tudo, desde a tag inicial do elemento até a tag final (inclusive)

```
<bookstore>
  <book category="CHILDREN">
    <title>Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WEB">
    <title>Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```

XML

- Um arquivo XML é definido por:
 - Um arquivo em formato ASCII
 - Tags aninhadas hierarquicamente (número indeterminado de sub-elementos)
 - Um único elemento (<tag>XPTO</tag>) raiz
 - Marcadores de início e fim de uma tag (diferente de html – por exemplo,
)
 - Um número indeterminado de instruções de processamento <?app ... ?> antes da raiz
 - Ex: <?xml ... ?>, <?xml-stylesheet ... ?>
 - Não faz parte do documento XML em si, mas serve para instruir aplicações específicas que utilizam o documento
 - Indeterminado número de atributos numa tag
 - <tag atr1="val1"... atrn="valn"></tag>

DTD - Document Type Definition

- Define os blocos de um documento XML.
- Define a estrutura do documento com uma lista de elementos e atributos possíveis
- Com um DTD, cada arquivo XML pode levar consigo uma descrição do seu próprio formato
 - Grupos independentes podem usar um padrão de troca de dados
 - A aplicação pode usar uma DTD padrão para verificar se os dados recebidos “de fora” são válidos (ou verificar seus próprios dados)

DTD: Exemplo

```
<!DOCTYPE root-element [element-declarations]>
<?xml version="1.0"?>
<!DOCTYPE recado [
  <!ELEMENT recado (para,de,assunto,corpo)>
  <!ELEMENT para (#PCDATA)>
  <!ELEMENT de (#PCDATA)>
  <!ELEMENT assunto (#PCDATA)>
  <!ELEMENT corpo (#PCDATA)>
]>
<recado>
  <para>João</para>
  <de>Maria</de>
  <assunto>Lembrete</assunto>
  <corpo>Não esqueça do chopp na sexta!</corpo>
</recado>
```

!DOCTYPE recado define o elemento raiz *recado*
!ELEMENT recado define *recado* contendo quatro elementos: "para, de, assunto, corpo"
!ELEMENT para define *para* do tipo "#PCDATA" etc.

Obs: PCDATA é texto que será analisado pelo computador (parser)

XSD - Esquema XML

- Documentos XSD (esquemas) definem uma gramática para documentos XML
 - Define os elementos que podem aparecer no documento (e seus tipos) e a relação entre eles (quem é filho de quem)
 - Evolução do DTD
- Características:
 - É definido na própria linguagem XML
 - É extensível
 - Suporta namespaces

XSD: Exemplo

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.w3schools.com"
xmlns="http://www.w3schools.com"
elementFormDefault="qualified">
```

```
<xs:element name="recado">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="para" type="xs:string"/>
      <xs:element name="de" type="xs:string"/>
      <xs:element name="assunto" type="xs:string"/>
      <xs:element name="corpo" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>

</xs:schema>
```

O elemento *recado* é um **tipo complexo** porque contém outros elementos. Os outros (to, from, heading, body) são **tipos simples** porque não contém outros elementos.

XML Namespace

- Os namespaces são utilizadas para distinguir os nomes dos elementos
- Alguns fabricantes utilizam urls que contém descrições do formato xml
 - Por exemplo, <http://www.w3.org/2001/XMLSchema>
- Além do atributo xmlns, o elemento raiz schema possui um atributo chamado targetNamespace, o qual permite a definição de um namespace
- O uso deste atributo permite que os elementos definidos neste esquema possam ser referenciados por outros documentos XML

XML

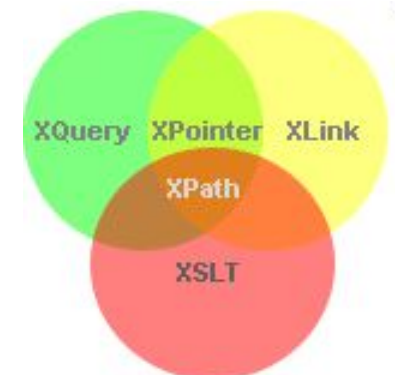
- Definições: XML, Namespace, DTD, XSD
- Vantagens: documento texto, formato extensível, reuso de ferramentas de manipulação
- Desvantagens: documento longo, manipulação manual ou através de bibliotecas
 - <http://xmlsucks.org/>

Visualizando XML

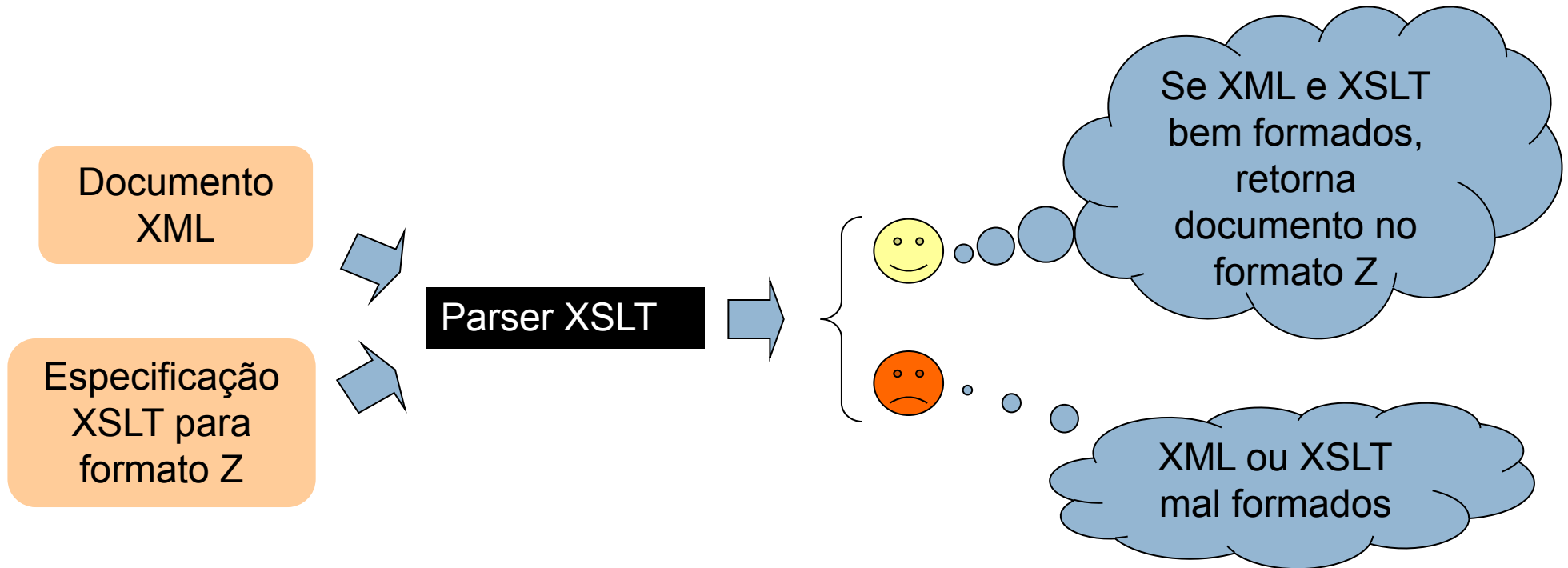
- Arquivos XML não geram uma forma visual como os arquivos HTML
 - Usa-se CSS, XSLT ou Javascript para processar e mostrar as informações no browser
- XSLT: transforma XML em outra coisa
 - HTML para mostrar no browser
 - Outro xml para transferir/comunicar com outras aplicações

XSLT

- ❑ XSLT significa *Extensible Stylesheet Language Transformation* (<http://www.w3.org/TR/xslt20/>)
- ❑ XSLT permite a transformação de um documento XML num documento em algum outro formato
- ❑ Também é definida utilizando o formato XML
- ❑ Utiliza a linguagem XPath para fazer referência aos elementos de um documento XML



XSLT



Exemplo

This is Raphael's "Foligno" Madonna, painted in 1511–1512.

```
<?xml version="1.0" encoding='UTF-8'?>
<painting>

<caption>This is Raphael's "Foligno" Madonna, painted
in <date>1511</date>–<date>1512</date>.
</caption>
</painting>
```



XML

- Usos:
 - Trocas de informações entre aplicativos
 - Arquivos de configuração
 - Formatos de entrada e/ou saída de aplicativos
 - Etc.
- Aplicações: XSLT, RSS, Web Services...

