

Introdução XML

Vanessa Braganholo
{vanessa@ic.uff.br}

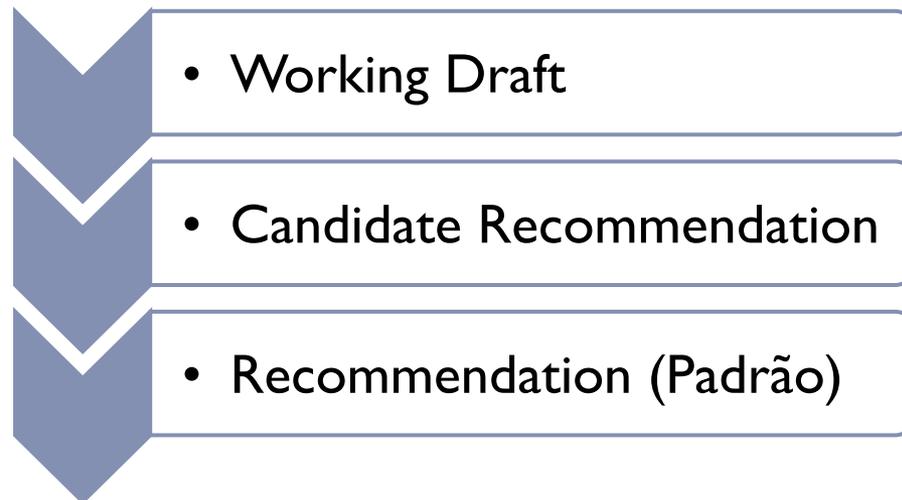
O que é XML?

- ▶ XML = eXtensible Markup Language
- ▶ Padrão para marcação de dados na Web, com foco na descrição do conteúdo



W3C – www.w3.org

- ▶ Órgão responsável pela padronização de iniciativas ligadas à Web
 - ▶ Ex.: HTML, XML e iniciativas relacionadas, entre outros
- ▶ Especificações dessas iniciativas são classificadas de acordo com seu nível de “maturidade”

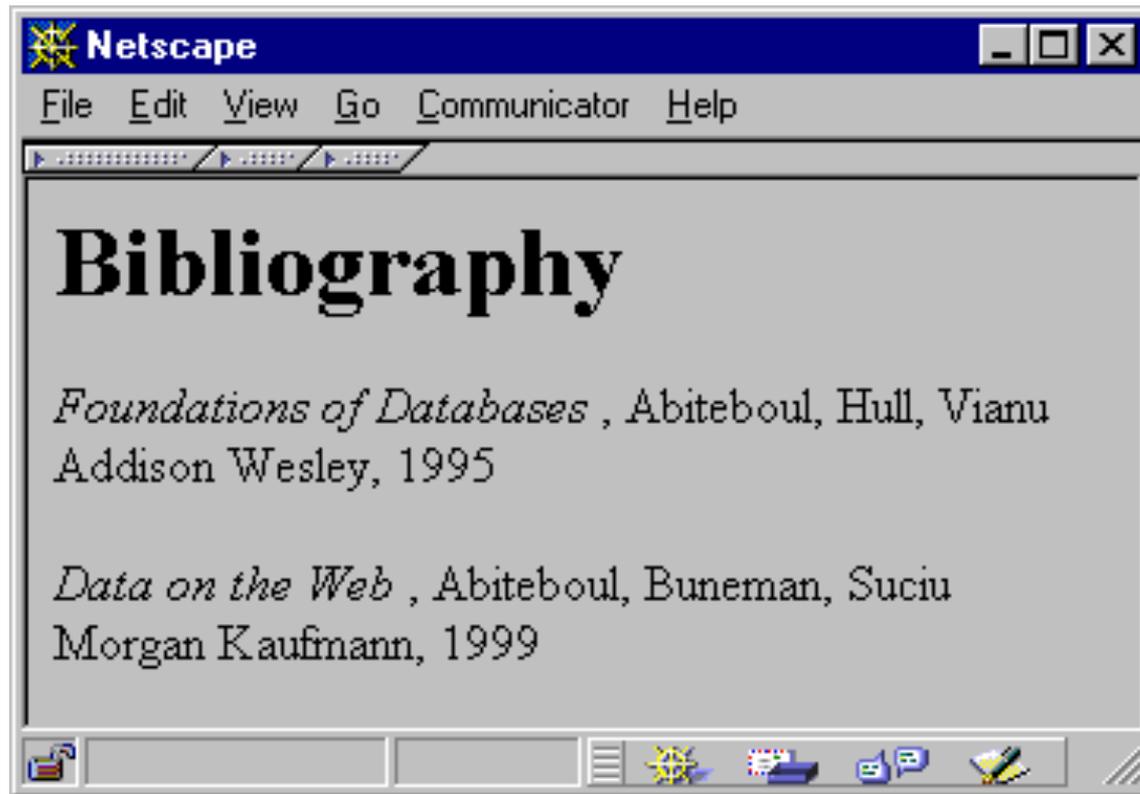


HTML x XML

- ▶ HTML – descreve o **formato** do documento
 - ▶ HTML tem um conjunto fixo de tags e não descreve conteúdo
- ▶ XML – descreve o **conteúdo** do documento
 - ▶ Usuário define suas próprias tags para criar uma estrutura
 - ▶ Um documento XML não tem nenhuma instrução para apresentação



De HTML para XML...



HTML descreve a **apresentação!**

Fonte HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML>
  <HEAD><TITLE>A bibliography on Databases</TITLE>
  <META content="text/html; charset=windows-1252" http-
equiv=Content-Type>
  <META content="MSHTML 5.00.2314.1000" name=GENERATOR>
</HEAD>
<BODY>
  <h1> Bibliography </h1>
  <p> <i> Foundations of Databases </i> Abiteboul, Hull, Vianu <br>
  Addison Wesley, 1995
  <p> <i> Data on the Web </i> Abiteoul, Buneman, Suciu <br>
  Morgan Kaufmann, 1993
</BODY>
</HTML>
```

HTML: Conjunto pré-definido de elementos (tags) para especificação das dimensões de estrutura e apresentação de um documento

Fonte XML

XML: Elementos (tags) definidos pelo usuário da linguagem e servindo para descrever o conteúdo e a estrutura.

```
<bibliography>
  <book>
    <title> Foundations... </title>
    <author> Abiteboul </author>
    <author> Hull </author>
    <author> Vianu </author>
    <publisher> Addison Wesley </
publisher>
    <year> 1995 </year>
  </book>
  ...
</bibliography>
```

XML descreve o conteúdo!!!



Sintaxe XML

XML (<http://www.w3.org/TR/xml/>)

▶ Documento XML

- ▶ Seqüência de elementos que englobam texto ou outros elementos
- ▶ Elementos podem conter atributos

XML

▶ Elemento

- ▶ Delimitado por marcas (*tags*)
- ▶ Possuem uma marca inicial e uma marca final
- ▶ Tudo o que estiver delimitado por essas marcas faz parte do conteúdo do elemento
- ▶ Ex: `<empregado>João</empregado>`

▶ Atributo

- ▶ Podem aparecer dentro da marca **inicial** de um elemento
- ▶ Ex: `<empregado cod="E01">João</empregado>`

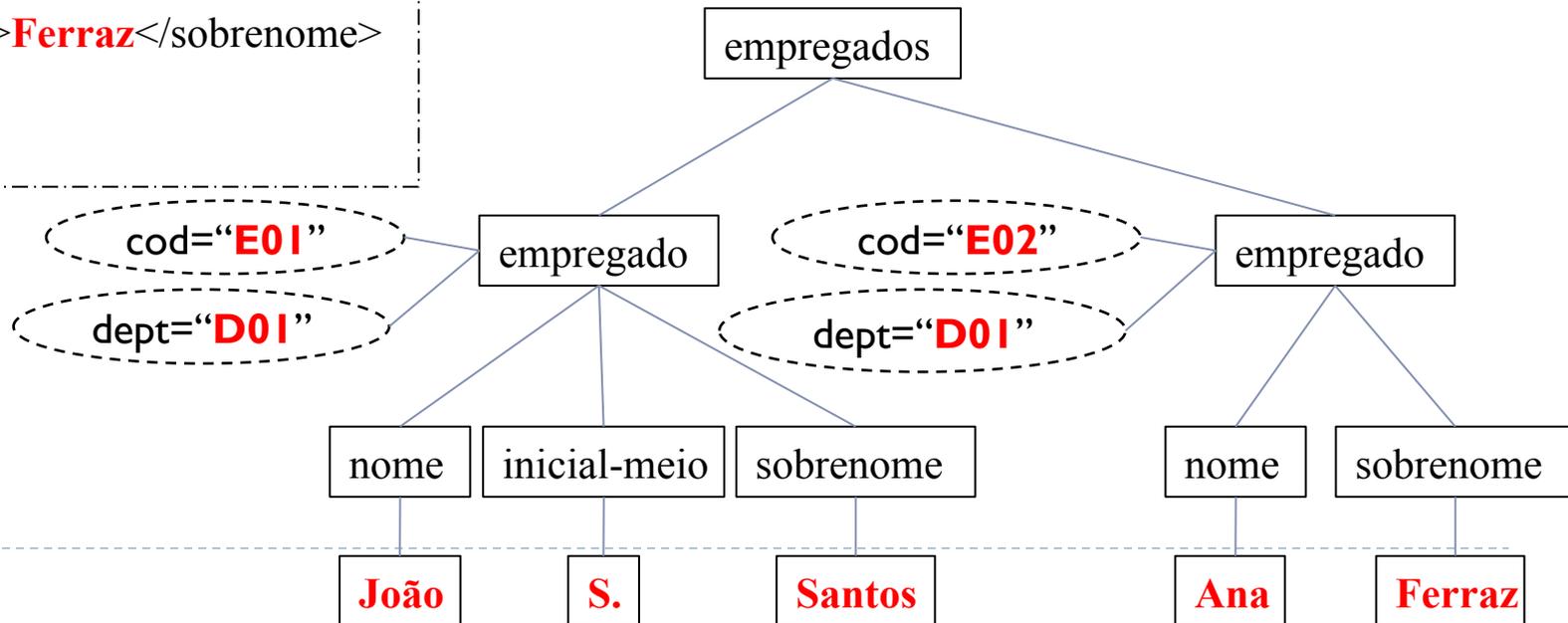
Exemplo de Documento XML

```
<? xml version="1.0" ?>
<empregados>
  <empregado cod="E01" dept="D01">
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
</empregados>
```

Documentos XML como árvores

```
<? xml version="1.0" ?>
<empregados>
  <empregado cod="E01" dept="D01">
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
</empregados>
```

- ▶ Elementos, atributos e **texto** → nodos
- ▶ Relações pai/filho → arestas



Documentos XML Bem-Formados

- ▶ Por representar uma estrutura de árvore, algumas restrições se aplicam a documentos XML
 - ▶ Raiz única (elemento documento ou elemento raiz)
 - ▶ Todas as marcas são fechadas
 - ▶ Elementos são bem aninhados (marcas fecham na ordem inversa à em que foram abertas)
 - ▶ Exemplo de elemento não bem-aninhado
`<empregado><nome>João</empregado></nome>`
 - ▶ Atributos não se repetem no mesmo elemento
 - ▶ Nomes de elementos são sensíveis a maiúsculas e minúsculas

Tipos de Elemento

▶ Composto

- ▶ Contém outros (sub)-elementos

```
<empregado>  
  <nome>Ana</nome>  
  <sobrenome>Ferraz</sobrenome>  
</empregado>
```

▶ Textual

- ▶ Contém somente texto

```
<nome>Ana</nome>
```

▶ Misto

- ▶ Contém texto e sub-elementos

```
<endereco> Rua das Flores, 75  
<cidade>Rio de Janeiro</cidade></  
endereco>
```

▶ Vazio

- ▶ Elemento sem conteúdo

```
<engenheiro></engenheiro>  
<engenheiro/>
```

Outras considerações importantes

- ▶ Elementos são ordenados
- ▶ Atributos não são ordenados

```
<? xml version="1.0" ?>
<empregados>
  <empregado cod="E01" dept="D01">
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
</empregados>
```

```
<? xml version="1.0" ?>
<empregados>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
  <empregado cod="E01" dept="D01">
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
</empregados>
```



Documentos Diferentes

Outras considerações importantes

- ▶ Elementos são ordenados
- ▶ Atributos não são ordenados

```
<? xml version="1.0" ?>
<empregados>
  <empregado dept="D01" cod="E01" >
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
</empregados>
```

```
<? xml version="1.0" ?>
<empregados>
  <empregado cod="E01" dept="D01">
    <nome>João</nome>
    <inicial-meio>S.</inicial-meio>
    <sobrenome>Santos</sobrenome>
  </empregado>
  <empregado cod="E02" dept="D01">
    <nome>Ana</nome>
    <sobrenome>Ferraz</sobrenome>
  </empregado>
</empregados>
```



Elementos x Atributos

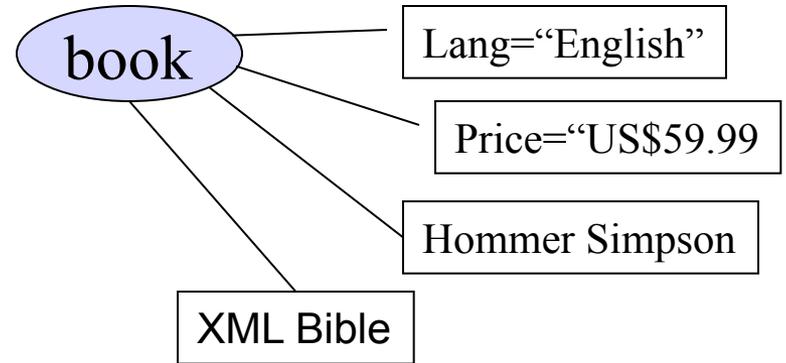
- ▶ Não há regras
- ▶ Atributos apresentam algumas restrições
 - ▶ Não são extensíveis
 - ▶ Não permitem múltiplos valores
 - ▶ Não descrevem estruturas
- ▶ Recomendação: em geral, preferir elementos, e usar atributos para informações secundárias
- ▶ Metadados (dados sobre os dados) devem ser representados como atributos

Ex: `<price currency="US">59.99</price>`

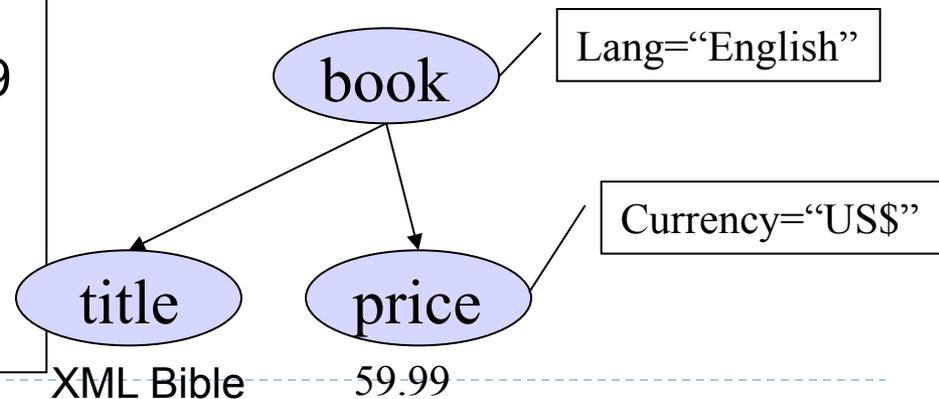


Elementos x Atributos

```
<book lang="English" price="US  
$59.99"  
title="XML Bible"  
author="Hommer Simpson">  
...  
</book>
```



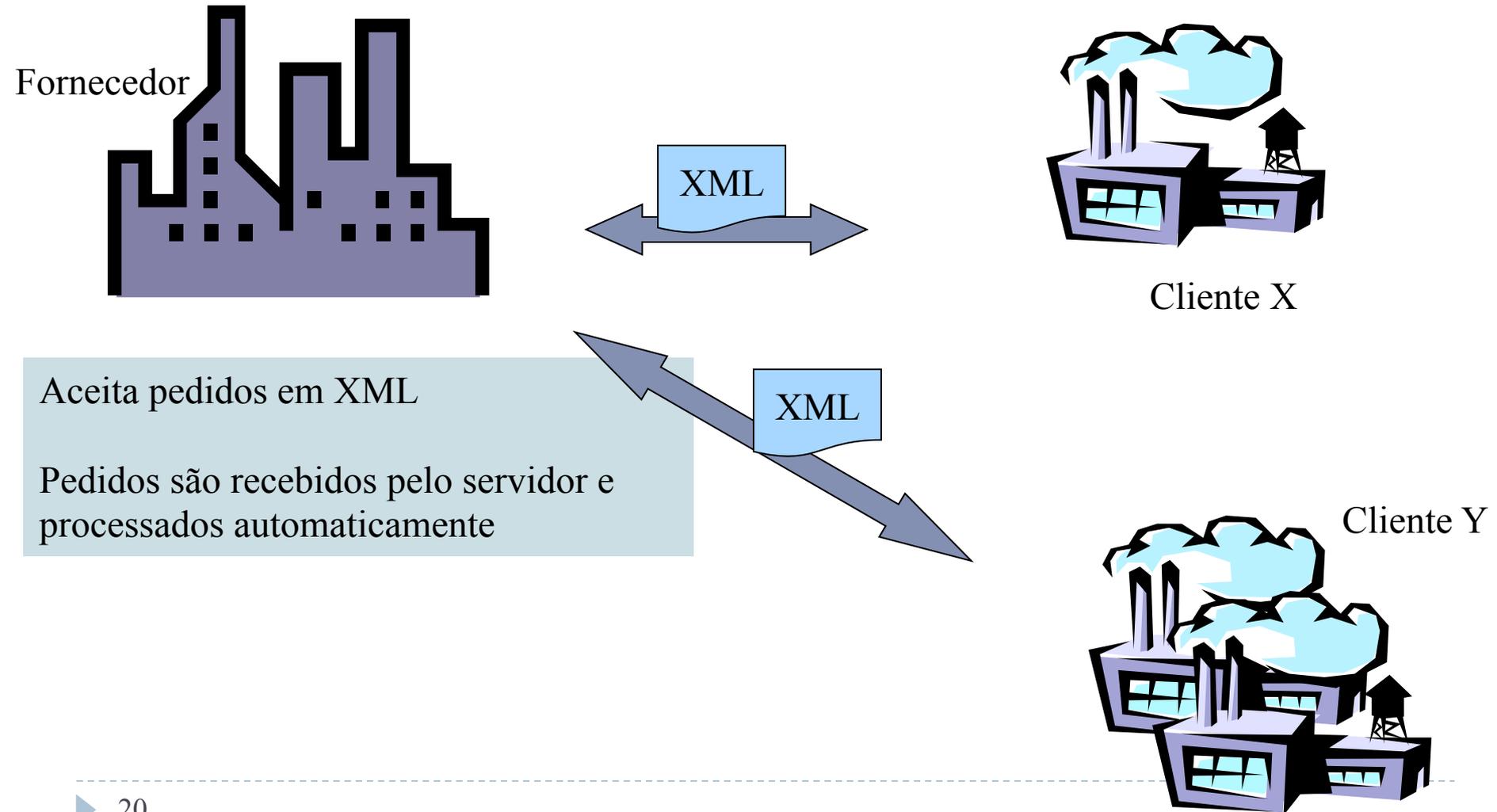
```
<book lang="English">  
  <price currency="US$"> 59.99  
</price>  
  <title>XML Bible </title>  
  ...  
</book>
```



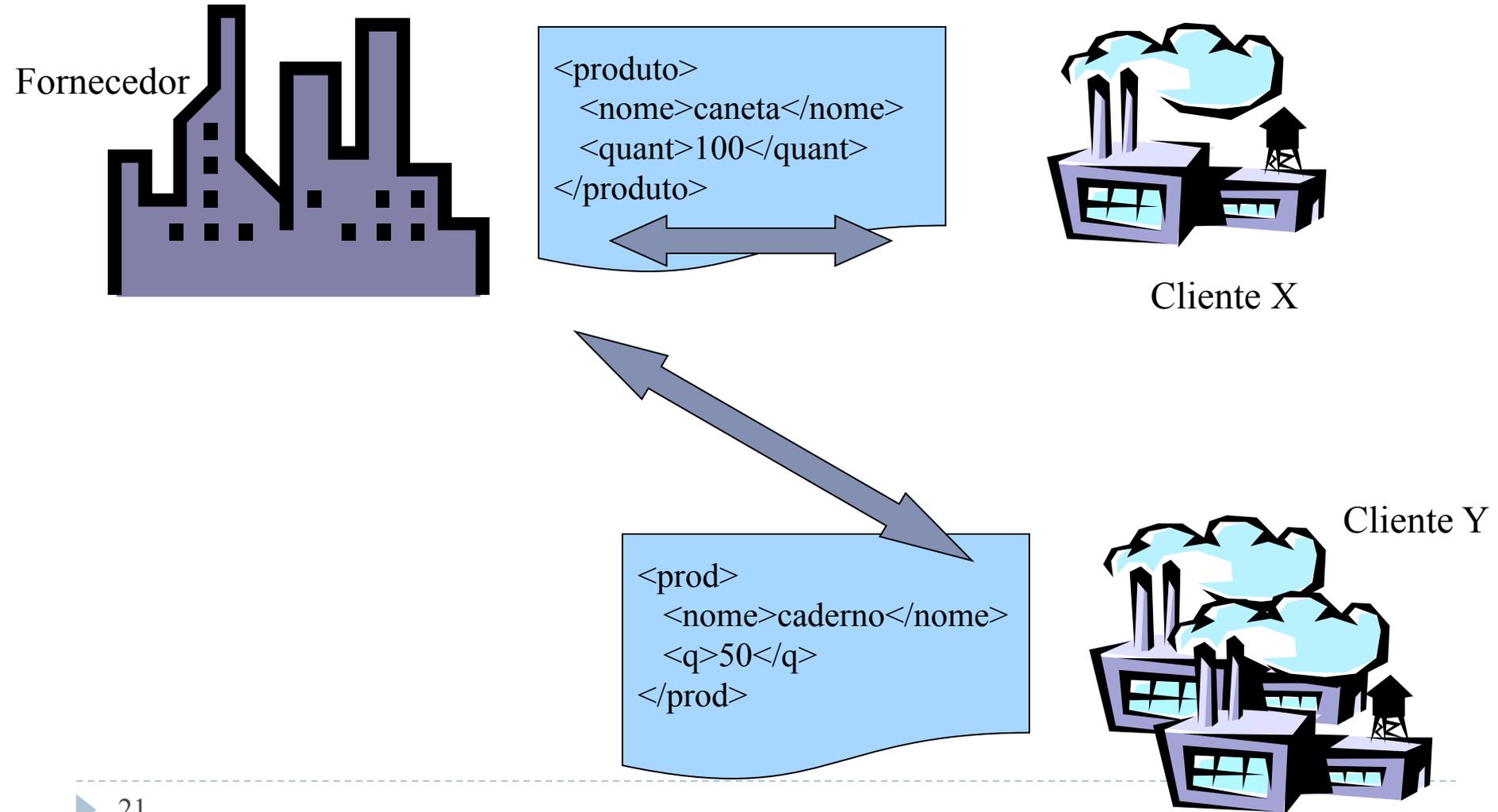
Interoperabilidade

- ▶ Chave do sucesso de XML:
 - ▶ Usuários podem definir suas próprias marcas
- ▶ E a interoperabilidade, como fica?
 - ▶ Vamos considerar um exemplo prático

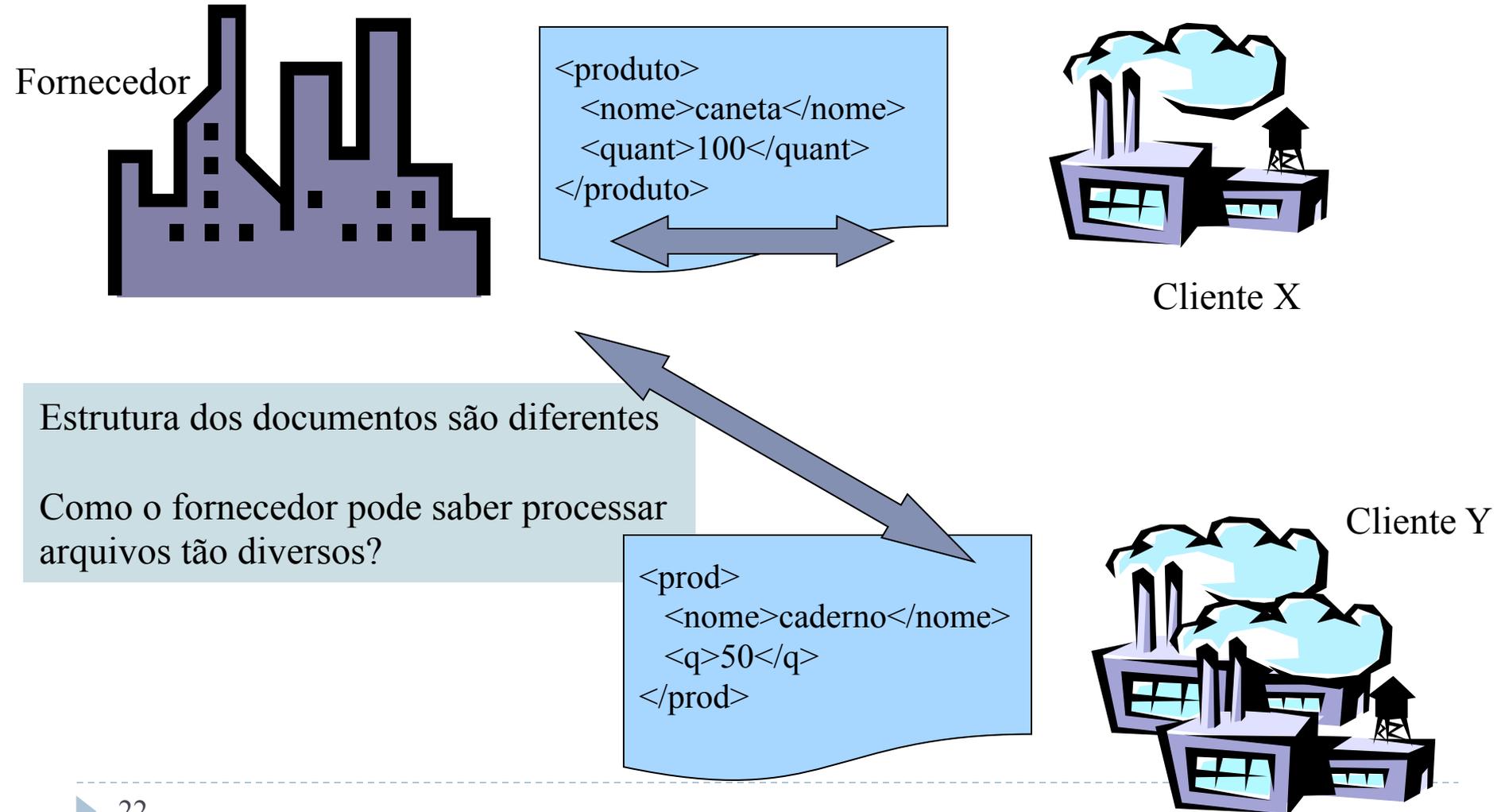
Exemplo prático: interoperabilidade



Exemplo prático: interoperabilidade



Exemplo prático: interoperabilidade



Solução

- ▶ Pode-se definir um vocabulário usando uma linguagem de esquemas para XML (DTD ou XML Schema)
 - ▶ Fornecedor define o vocabulário (estrutura, nomes das marcas, tipos de dados)
 - ▶ Cliente usa o vocabulário para enviar os pedidos

Interoperabilidade

