

O GUIA PRÁTICO

DA

XML

Conceito, exemplos, prática e aplicações da linguagem universal.



Áreas de Uso, Exemplos, Normas, Elementos, Atributos, Sintaxe, DTDs, Schemas, Folhas de Estilo, Parsers, DOM, SAX, Aplicações, Processamento de Dados, Vocabulários e Dialectos, Comércio Electrónico.

Paulo Heitlinger

Guia prático da XML

Conceito, exemplos,
prática e aplicações da
linguagem universal.

Áreas de Uso, Exemplos, Normas,
Elementos, Atributos, Sintaxe,
DTDs, Schemas, Folhas de Estilo,
Parsers, DOM, SAX, Aplicações,
Processamento de Dados,
Vocabulários e Dialectos,
Comércio Electrónico.

XML é o **formato universal para partilha de dados entre aplicações**. O conceito XML é simples e as possibilidades são inúmeras. Documentos em formato XML podem conter: Bases de Dados, Transacções Comerciais, Catálogos de Produtos, Gráficos Vectoriais, Equações Matemáticas, Fórmulas Químicas, Relatórios Financeiros, Dados Bibliográficos, Anúncios Publicitários, enfim, quase todos os dados estruturados, em documento de texto. **Paulo Heitlinger**.

Reservados todos os direitos por Centro Atlântico, Lda.
Qualquer reprodução, incluindo fotocópia, só pode ser feita
com autorização expressa dos editores da obra.

O GUIA PRÁTICO DA XML

Colecção: Tecnologias

Autor: **Paulo Heitlinger**

Direcção gráfica: Centro Atlântico

Revisão: Centro Atlântico

Capa: Paulo Buchinho

© **Centro Atlântico, Lda., 2001**

Ap. 413 – 4764-901 V. N. Famalicão

Porto - Lisboa

Portugal

Tel. 808 20 22 21

geral@centroatlantico.pt

www.centroatlantico.pt

Fotolitos: Centro Atlântico

Impressão e acabamento: Inova

1ª edição: Outubro de 2001

ISBN: 972-8426-45-3

Depósito legal: 171.888/01

Marcas registadas: todos os termos mencionados neste livro conhecidos como sendo marcas registadas de produtos e serviços, foram apropriadamente capitalizados. A utilização de um termo neste livro não deve ser encarada como afectando a validade de alguma marca registada de produto ou serviço.

O Editor e o Autor não se responsabilizam por possíveis danos morais ou físicos causados pelas instruções contidas no livro nem por endereços Internet que não correspondam às *Home-Pages* pretendidas.

Índice de Conteúdo

1. XML, os conceitos básicos	9
1.1. Introdução.....	10
Para quem é destinado este livro?.....	10
Conhecimentos prévios	10
Agradecimentos, dedicatória.	11
1.2. Para que serve a XML?	12
"Língua-mãe" de novas linguagens de marcação.	14
1.3. O que é, precisamente, a "XML"?	16
"Dados autodescritíveis" – um primeiro exemplo.....	16
Dados estruturados	18
SGML, XML, HTML e XHTML	18
1.4. Editar e ver documentos XML	20
1.5. As diferenças essenciais: XML vs. HTML	22
XML não será a futura versão da HTML... ..	22
O "ponto forte" da XML: marcadores semânticos.	23
A extensibilidade dos elementos.	25
Em continuação.....	27
2. Bem formar: A sintaxe da XML	29
2.1. Exemplo de documento bem formado	30
As componentes de um documento XML	30
O que é, precisamente, um <elemento> XML?	32
2.2. Regras para o uso de marcadores.....	33
Elementos sem conteúdo.	33
Espaços em branco.	34
Convenções para a denominação de <elementos>	34
O Document Object Model: Estrutura em árvore	35
2.3. Elementos com atributos	38
Preferir atributos ou elementos?	40

3.	DTDs: Fixar modelos de conteúdo	41
3.1.	Para quê convenções sobre documentos XML?	42
	Cenário para aplicações práticas de XML com DTDs	44
3.2.	Declarar Tipos de Documentos	46
3.3.	Tipos de declaração	47
	Declarações de Tipos de Elementos	47
	Declaração DOCTYPE	48
	Declaração de elementos	48
	Declarações para o valor dos elementos	48
	Declaração de <sub-elementos>	50
	Declaração de atributos (ATTLIST Definition)	51
3.4.	Links para DTDs	54
4.	XML-Schema	55
4.1.	Os limites das DTD convencionais	56
4.2.	Os potenciais da XML-Schema	57
4.3.	A validação de documentos XML com schemas	59
4.4.	A construção de um schema	60
4.5.	Páginas Web sobre XML-Schema	64
5.	XLink: Apontar para recursos	65
5.1.	A utilidade do XLink	66
	As componentes do XLink	66
	Link simples	66
	Links estendidos com xlink:type="extended"	66
	Web-sites de utilidade	68
6.	Namespaces	69
6.1.	O conceito dos espaços de nome	70
7.	Analisar documentos XML	71
7.1.	Parsers XML	72
	Parsers tipo DOM	72
	Parsers tipo SAX: "event driven processing"	74
7.2.	Editores e Parsers	75
	MSXML: o Parser da Microsoft	75
7.3.	Outros Parsers e ferramentas XML	76

8. Publicar XML com folhas de estilo	79
8.1. Formatar documentos XML com CSS ou XSL	80
Cascading Style Sheets	80
8.2. Formatar documentos XML com XSL	82
XSL ou CSS? Prós e contras.	82
Cross Media Publishing	83
8.3. As três componentes da XSL	85
XSLT, XPath, XSL FO	85
8.4. XSLT: XSL-Transformations	86
Templates: os padrões XSL	86
O uso dos templates.	87
8.5. XSL: Alguns exercícios práticos	88
XML+XSL ->HTML	88
Style Sheet de tipo XSL para gerar output HTML	89
Desempenho duma transformação XSL com XMLWriter	92
Alternativa: Usar o XSL Engine do Internet Explorer	93
Melhor ainda: XSLT "on the server"	94
9. Processar dados XML	97
9.1. MS Access: Importar e exportar XML	98
9.2. StarOffice da SUN	100
9.3. Navegar dados XML com JavaScript	101
9.4. Desenvolver aplicações: Java ou .NET?	105
A estratégia da SUN: »Portable Code, Portable Data«	105
A ofensiva .NET e a plataforma BizTalk da Microsoft	106
10. Dialectos XML	107
10.1. Descendentes da XML: vocabulários e dialectos	108
10.2. MathML: Codificar expressões matemáticas	109
Codificação de conteúdo vs. codificação de apresentação.	110
10.3. CML - Chemical Markup Language	115
10.4. UDDI - Universal Description, Discovery and Integration	116
10.5. WAP e WML	118
10.6. SVG - Gráficos vectoriais	120
10.7. Outros dialectos e projectos XML: uma pequena selecção	127

11. XML no Comércio Electrónico	131
11.1 EDI, o precursor da XML	132
Intercâmbio de informação via EDI	132
Os benefícios do EDI	134
Os problemas do EDI	134
Os vocabulários XML para o Comércio Electrónico	135
11.2. eCommerceXML e XML/EDI	136
11.3. Business Integrator	137
11.4. BizTalk Server 2000 da Microsoft	138
11.5. Tamino XML DB da Software AG	139
11.6. Gestão de Conteúdos com o Dialog Server	140
12. XML online: Links de utilidade	141
Tutoriais.....	142
Editores, Parsers, Aplicações	142
Comércio Electrónico	142
Empresas, Entidades	142
Directórios	143
Newsgroups na Usenet	143
13. Glossário	145

1. XML, os conceitos básicos

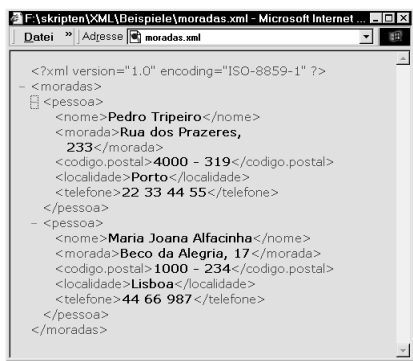
Neste capítulo vamos editar um documento, guardá-lo como ficheiro XML, e visualizá-lo com um browser. Verificamos logo que não existe uma formatação satisfatória; mas descobrimos o "ponto forte" da XML: Com <marcadores> tipo XML podemos não só descrever o conteúdo, como também fixar a estrutura lógica desse mesmo conteúdo.

A flexibilidade da XML provém da faculdade de transportar variadíssimos tipos de dados e de mantê-los estruturalmente coesos. A XML serve magnificamente para a estruturação de qualquer tipo de dados e para descrevê-los sem dúvidas ou ambiguidades em formato de texto.

Isto porque a XML permite definir a linguagem de marcação mais adequada a um dado tipo de documentos.

Documentos em formato XML podem conter, por exemplo,

*Transacções Comerciais,
Catálogos de Produtos,
Gráficos Vectoriais,
Equações Matemáticas,
Anúncios Publicitários, enfim:
todos os dados que necessitem
de uma representação
estruturada...*



```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<moradas>
  <peessoa>
    <nome>Pedro Tripeiro</nome>
    <morada>Rua dos Prazeres,
      233</morada>
    <codigo.postal>4000 - 319</codigo.postal>
    <localidade>Porto</localidade>
    <telefone>22 33 44 55</telefone>
  </peessoa>
  <peessoa>
    <nome>Maria Joana Alfacinha</nome>
    <morada>Beco da Alegria, 17</morada>
    <codigo.postal>1000 - 234</codigo.postal>
    <localidade>Lisboa</localidade>
    <telefone>44 66 987</telefone>
  </peessoa>
</moradas>
```

1.1. Introdução

Para quem é destinado este livro?

Este livro foi escrito para todas as pessoas que querem uma apresentação concisa – em língua portuguesa – da nova tecnologia de intercâmbio de dados abreviada por "XML".

Com este livro pretendemos transmitir os conhecimentos necessários para uma **apreciação realista** – não só do presente uso, como também das futuras potencialidades da **Extensible Markup Language**. Explica-se o conceito, os módulos, apresentam-se vários vocabulários derivados da XML e descrevem-se algumas aplicações que trabalham com esta tecnologia de informação.

Paralelamente, este livro fornece uma iniciação do tipo "aprender fazendo" (*learning by doing*) em forma de **exercícios práticos** – para pessoas que pensem estender as suas actividades profissionais a este novo sector.

Conhecimentos prévios.

Mas este livro não foi escrito apenas para profissionais; qualquer pessoa com conhecimentos básicos de Informática e de computadores pessoais poderá seguir os temas aqui abordados. Contudo, terá alguma vantagem se já possuir alguns conhecimentos – ou até alguma prática – nos seguintes campos:

- HTML e CSS, Internet em geral,
- Bases de Dados,
- Linguagens de scripting – como, p.ex., JavaScript.

Se quiser fazer os exercícios práticos incluídos neste livro, necessitará de um PC – trabalhando preferivelmente em ambiente *Windows 95, 98, 2000, XP* ou *Win NT* –, do *browser MS Internet Explorer versão 5* ou superior, e de um *Editor de Texto* simples.

Uma ligação à Internet será útil para todos os que desejem...

... fazer o download de diversas aplicações e/ou exemplos mencionados neste livro,

... completar os seus conhecimentos sobre os temas abordados neste livro consultando os inúmeros Tutoriais, White Papers e páginas Web.

Informações de grande utilidade, oferecidas por empresas de software e programadores *freelance* – principalmente nos idiomas inglês, alemão, francês e castelhano. (A informação em português é escassa.)

- ... participar nos diversos Grupos de Discussão (newsgroups) da Usenet, que servem para esclarecer e discutir a presente evolução da XML – uma tecnologia ainda bem nova!
- ... para os leitores que, por e-mail, queiram entrar em contacto com o autor ou com a Editora para tirar qualquer dúvida, ou que queiram contribuir com aplausos, críticas e sugestões para o melhoramento das próximas edições deste livro.

Agradecimentos, dedicatória.

Tenho a agradecer primeiro a **Libório Manuel Silva**, Director da Editora Centro Atlântico, pela rápida reacção positiva quando lhe propuz em Agosto de 2001 a publicação deste livro. A **Isabel Canhola**, amiga de longa data e gerente da **Ponto i Lda.**, tenho a agradecer, entre outras facilidades, o espaço de trabalho que foi necessário para a redacção deste livro.

A **Mariana Figueroa** e **Ricardo Faria** agradeço a correcção do manuscrito e vários melhoramentos de estilo. A **Spike Lee** e a **Birgit Wegemann** agradeço o imperativo sempre actual: “Do the right thing!”

Este livro, paginado com PageMaker 6.5, e tipografado com a *Rotis* do mestre Otl Aicher, está dedicado a **Ted Nelson**, Planeta Xanadu:

“Tinhas razão, Ted!”

Porto, Outubro de 2001

Paulo Heitlinger

1.2. Para que serve a XML?

Desde a sua especificação em 1998, a **eXtended Markup Language** (XML) rapidamente adquiriu grande popularidade entre os peritos informáticos. Embora o **World Wide Web Consortium** (W3C) tenha fixado a versão 1.0 da XML há apenas três anos, vemos hoje este conceito já firmemente integrado no **Comércio Electrónico**. E não só no *Electronic Business*; vemos dialectos XML avançar céleres nos mais variados sectores: Medicina, Química, Biologia, Matemática, Ensino e Formação, Publicidade, Comunicação Social, Artes Gráficas, Desktop Publishing, Bibliotecas e Arquivos, etc.

Ao que se deve este sucesso tão rápido?

Um factor determinante para o rápido avanço da XML é o suporte garantido por empresas líderes do mercado do software – empresas como a Microsoft, Oracle, SAP, Software AG, IBM, SUN – entre outras de menor importância...

Vejamos alguns exemplos actuais: A XML é uma tecnologia fundamental da nova plataforma **.NET** da Microsoft, que irá permitir «a criação de serviços Web distribuídos que possam interoperar em ambientes heterogéneos». Quando do lançamento da iniciativa **BizTalk**, Bill Gates afirmou: «para seduzir mais de um milhão de novas empresas para o Comércio Electrónico, temos que ajudar as suas aplicações a falar a linguagem do negócio de forma homogénea. Até agora, era muito difícil para as empresas fazerem negócio com base na Internet, devido à falta de um vocabulário técnico único para descrever os processos funcionais».

Para a **PeopleSoft**, a implementação da XML como tecnologia de partilha de dados serve para a extensão da sua solução ERP para contemplar aplicações orientadas para o Comércio Electrónico.

A **Software AG** é a primeira grande empresa de software europeia a apostar completamente e totalmente no conceito XML.

Formato universal para partilha de dados.

Uma outra determinante do êxito da XML está fundamentada na sua própria “natureza”: o formato XML tornou-se, em pouco tempo, um **formato universal para a partilha de dados entre aplicações**. O conceito e a filosofia da XML são simples, os seus derivados e as possíveis aplicações são inúmeras. Documentos de formato XML podem conter, por exemplo:

- Transacções Comerciais,
- Catálogos de Produtos,
- Relatórios Financeiros ou Estatísticos

- Grafismos Vectoriais,
- Equações Matemáticas,
- Anúncios publicitários,
- Todos os dados que necessitem de uma representação estruturada.

Standard internacional.

Onde antes existiam bastiões proprietários, a nova *língua franca* trouxe uma grande simplificação para a transferência e para a partilha de dados. Com a **meta-linguagem XML** facilmente se definem **linguagens de marcação** para a partilha de informação entre aplicações baseadas em sistemas incompatíveis. Incompatíveis, por trabalhar com formatos proprietários.

Convém salientar de antemão que a partilha de dados muito estruturados entre computadores heterogéneos figura entre as aplicações principais da XML. »A aplicação mais importante da XML nos próximos anos será a comunicação de computador a computador, ou seja, a partilha de dados – e não o diálogo entre o homem e o computador«, afirmou com razão Simon Phipps, da IBM.

A integração fácil entre aplicações de uma empresa – ou de empresas diferentes – está entre as grandes potencialidades da XML.



Para o que a XML não serve...

Se bem que persista a tendência de associar e de comparar a XML com a função delegada à HTML, tratam-se de dois conceitos com fins e propósitos bem diferentes. É por isso que a XML não será a "herdeira" ou a "continuação" da HTML.

Vejamos porque não. A HTML é a linguagem de **formatação** utilizada para definir e integrar os elementos de um documento e para fornecer um **layout** específico a páginas Web compostas de texto, imagens e, possivelmente, de elementos multimédia. Para esse tipo de formatação, a XML precisa sempre da assistência da XSL – uma Style Sheet Language bastante mais complexa e difícil de dominar que a CSS da HTML.

Não cremos pois que a XML vá brevemente definir "o futuro da WWW". Nem tão pouco cremos que "a XML vai pôr a HTML de lado".

Para efectivar *layouts* e realizar um bom *Screen Design* de páginas Web, a HTML continuará a ser a **ferramenta de eleição**. Em conjunto com as Cascading Style Sheets, a HTML tem – em termos de desenho gráfico – nítidas vantagens sobre a XML. Para os amadores é fácil de aprender, e nas mãos de um profissional é funcional e efectiva (desde que o dito profissional se absteinha de usar alguns conhecidos 'editores' de HTML!).

A XML terá sim um forte impacto na futura evolução do World Wide Web e na HTML, mas não será a substituta da HTML.

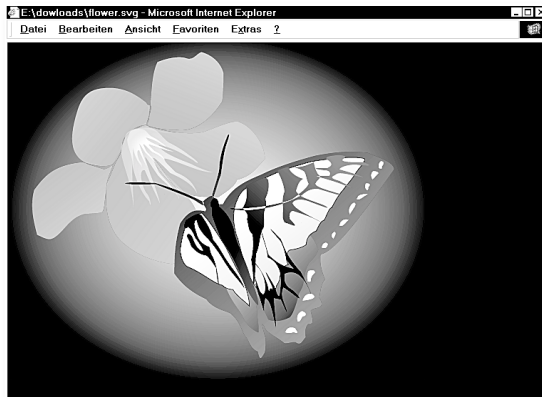
"Língua-mãe" de novas linguagens de marcação

A XML é a "língua-mãe" de outras linguagens. Em só três anos, a XML demonstrou já um potencial muito frutífero, pois serviu para gerir um sem número de linguagens de marcação.

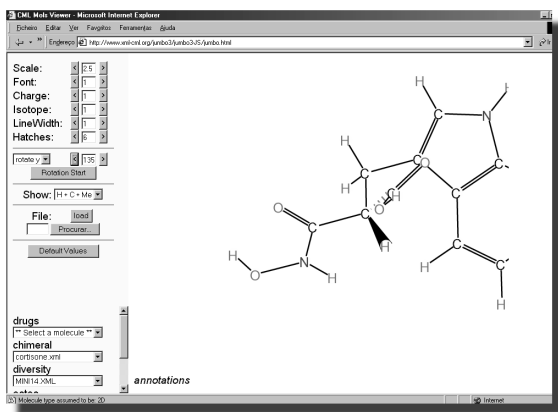
Entre os mais conhecidos dialectos (também chamados "vocabulários") derivados da XML figuram MathML, SVG, AdXML, ICE, SMIL, SOAP, BML, Biz-Talk, UDDI, WML, e muitos mais.

Visualização de dados.

A XML vai fomentar o surgimento de uma nova geração de aplicações de manipulação e visualização de dados. Hoje existem poucos *browsers* capazes de visualizar documentos criados com dialectos XML. Esperamos que em breve surjam *browsers* mais universais, mais versáteis – de modo a permitirem, por exemplo, a apresentação de fórmulas complexas codificadas com MathML – um dialecto da XML próprio para a definição de fórmulas matemáticas.



Uma das mais surpreendentes aplicações da XML é a Linguagem de Marcação SVG, que permite realizar gráficos vectoriais desta qualidade. Contudo, é necessário um PlugIn para podermos ver o gráfico dentro da janela de um browser...



Uma outra espantosa aplicação da XML é a Linguagem de Marcação CML, que permite descrever e visualizar moléculas químicas complexas.

Contudo, também neste caso é necessário um PlugIn para ver e mover a molécula em 2D ou 3D dentro da janela do Browser MS Internet Explorer...

Standards e "Standards".

Nem tudo é um mar de rosas nos campos da XML. Como já observámos em outros meios, também aqui vemos a ingerência de grandes empresas de software – nem é necessário mencionar nomes – na tentativa de introduzir os seus conceitos proprietários, quase sempre em prejuízo dos standards internacionais. Mas nem sempre esta pressão parece ser nociva; no caso dos *XLM-Schemas* verificamos que este conceito está a "virar" standard internacional – para o contentamento quase geral dos técnicos empenhados na evolução da XML.

O aspecto mais interessante da XML é, paradoxalmente, também o mais problemático: A enorme profusão de "dialectos" e "vocabulários XML" demonstra a versatilidade desta meta-linguagem. Mas quais destes "dialectos XML" serão verdadeiros padrões? E quais são os standards que, por carência de adesão geral, já hoje são de facto standards puramente fictícios?

1.3. O que é, precisamente, a "XML"?

A Extensible Markup Language (XML) provê um conceito para descrever, armazenar, intercambiar e manipular dados estruturados.

Num documento de formato XML os dados são descritos em **texto** (Unicode). Sabemos bem que um formato de texto não é a solução mais económica para armazenar dados, mas muitas vezes é a mais eficaz. Mais eficaz, porque é mais fácil trabalhar com um texto, do que trabalhar com os formatos binários típicos de muitas bases de dados.

"Dados autodescritíveis" – um primeiro exemplo

Com a XML definem-se facilmente **linguagens de marcação** (Markup Languages). Usar XML é um modo muito flexível de criar "dados autodescritíveis" e partilhar esses dados via Internet, por Intranets, ou por qualquer outro modo de transporte.

A XML facilita **declarações precisas** – não só dos conteúdos de um documento, como também dos elementos convenientes à estruturação desses conteúdos.

Para demonstrar esta qualidade básica da XML, basta dar um vista de olhos a um exemplo bem simples, como o seguinte extracto de um documento XML:

```
....  
<peessoa>  
  <nome>Pedro Tripeiro</nome>  
  <morada>Rua dos Prazeres, 233</morada>  
  <codigo.postal>4000 - 319</codigo.postal>  
  <localidade>Porto</localidade>  
  <telefone>22 33 44 55</telefone>  
</peessoa>  
....
```

Este pequeno exemplo demonstra de forma evidente que a XML serve magnificamente para a estruturação de dados e para descrevê-los sem dúvidas ou ambiguidades em forma de um texto. Isto porque a XML permitiu ao autor do documento definir os seus próprios <elementos>, criando e usando os <marcadores> (=Tags) mais adequados à descrição e à estruturação dos dados em causa: nome, morada, código postal, etc.

Para todos os efeitos práticos, definimos já a nossa "MML" – a *Moradas Markup Language*, que nos vai servir para armazenar e estruturar informação do tipo "moradas"...