

Processamento Estruturado de Documentos

DOSSIER DA DISCIPLINA

José Carlos Leite Ramalho

9 de Setembro de 2002

1 Modalidade lectiva

Disciplina de opção oferecida aos quintos anos das Licenciaturas em Engenharia de Sistemas e Matemática e Ciências da Computação.

2 Objectivos

Familiarização e Utilização da tecnologia baseada em XML e XSL para suporte da publicação electrónica e como plataforma de intercâmbio entre diferentes sistemas e aplicações.

No fim do módulo, o aluno deverá ser capaz de montar uma plataforma para processamento documental usando XML, XSL e ferramentas derivadas, suportando todas as fases do ciclo de vida documental.

3 Conteúdo Programático

Durante o módulo serão abordados vários temas que se desenvolvem nas seguintes secções.

3.1 Introdução

Um pouco de história: SGML, HTML, XML, SML.

As origens destas normas. Como é que elas se inter-relacionam.

3.2 Documentação Estruturada e Anotação

O conceito de documento estruturado.

Os vários tipos de anotação e sua evolução nos últimos tempos.

As vantagens e as desvantagens da Anotação Descritiva.

Como é composta uma linguagem de anotação e qual a sua aplicação.

O conceito de meta-linguagem de anotação: o SGML e o XML.

3.3 Documentos XML: estrutura e conceitos

As componentes dum documento XML: a declaração, o DTD e a instância.

Documento válido versus documento bem-estruturado.

Elementos estruturais dum documento XML: elementos, atributos e entidades.

Alguns exemplos de documentos XML: o soneto, os sumários, algumas obras literárias e informação importada de bases de dados.

3.4 Ciclo de vida documental

Técnicas de Análise documental: diagramas, elm-trees.

Edição de documentos XML: utilização de editores estruturados.

Validação de documentos: integrada em editores ou outras ferramentas e na linha de comando.

Introdução à transformação e formatação de documentos.

Como armazenar documentos estruturados: bases de dados, LDAP, UDDI.

3.5 Desenvolvimento de DTDs e Schemas

Construção de DTDs e sua utilização no ciclo de vida documental.

Princípios básicos a serem observados no desenvolvimento de DTDs. Algumas técnicas.

Introdução aos Schemas: DTDs versus Schemas.

Comparação de Schemas com gramáticas independentes de contexto.

Desenvolvimento de Schemas para alguns casos de estudo.

3.6 XSL: estrutura e conceitos

O modelo de dados subjacente a um documento XML.

Transformação de documentos como transformação de árvores abstractas.

O XSL e a transformação declarativa de documentos XML.

Estrutura duma especificação em XSL e respectivo modelo de processamento.

Os vários operadores e funções do XSL.

3.7 Processamento imperativo de documentos

Utilização de linguagens de programação imperativas para transformar documentos XML.

Os dois modelos de programação: "tree-driven" (DOM), "event-driven" (SAX).
Alguns exemplos com aplicação: XML::DT, Omnimark.

3.8 XML: últimos desenvolvimentos

Algumas normas que irão interagir com o XML: XPointer, XLink, Namespaces, XPipe, MDTs.

3.9 XML e o intercâmbio da informação

Introdução ao XMI. Alguns exemplos de aplicação.

3.10 Algumas aplicações XML

- EAD: Encoded Archive Description.
- DocBook: Scientific Documentation.
- HL7: Medical Records.
- MathML: Mathematical Markup Language.
- XCSL: XML Constraint Specification Language.
- XTM: XML Topic Maps.
- WSDL: XML Web Services.
- SVG: Standard Vector Graphics

4 Avaliação

Vão existir dois elementos de avaliação, o exame individual e um projecto a desenvolver em equipe de 2 elementos, *durante o período lectivo*.

O aluno poderá realizar ambos ou apenas um deles sendo a nota final igual à nota máxima obtida.

Referências

- [HM01] Elliot Rusty Harold and W. Scott Means. *XML in a Nutshell*. O'Reilly, 2001.
- [MA96] Eve Maler and Jeanne Andaloussi. *Developing SGML DTDs: From Text to Model to Markup*. Prentice-Hall, 1996.
- [Ram00] José Carlos Leite Ramalho. *Anotação Estrutural de Documentos e sua Semântica*. PhD thesis, Escola de Engenharia - Universidade do Minho, Julho 2000.
- [RH02] José Carlos Ramalho and Pedro Rangel Henriques. *XML e XSL: da teoria à prática*. Tecnologias da Informação. FCA, 2002.
- [Tid01] Doug Tidwell. *XSLT*. O'Reilly, August 2001.
- [Wil00] Kevin Williams. *Professional XML Databases*. Wrox Press, 2000.
- [Wil01] Heather Williamson. *XML: the complete reference*. Osborne/McGraw-Hill, 2001.