DTD Nguyễn Hồng Phương Email: phuong.nguyenhong@hust.vn Site: http://is.hut.edu.vn/~phuongnh Bộ môn Hệ thống thông tin Viện Công nghệ thông tin và Truyền thông Đại học Bách Khoa Hà Nội

Nội dung DTD là gì? Khai báo DTD nội Khai báo DTD ngoại Khai báo phần tử Khai báo thuộc tính Phần tử vs. Thuộc tính Thực thể

DTD là gì? ☐ Document Type Definition ☐ Là tài liệu dùng để định nghĩa kiểu dữ liệu cho các phần tử của tài liệu XML. ☐ Khi đọc tài liệu XML, chỉ cần đọc phần DTD là sẽ biết được cấu trúc XML

Tại sao sử dụng DTD? Với một DTD, mỗi file XML có thể mang một mô tả định dạng của nó. Các nhóm độc lập có thể chấp nhận một chuẩn DTD để trao đổi dữ liệu Một ứng dụng có thể sử dụng một DTD chuẩn thể kiểm tra dữ liệu nhận về từ bên ngoài có hợp lệ hay không.

Khai báo DTD nội Nếu DTD được khai báo bên trong file XML, nó nằm trong định nghĩa DOCTYPE <!DOCTYPE root-element [element-declarations]>



Giải thích

- □ !DOCTYPE note định nghĩa phần tử gốc của tài liêu là note
- ☐ !ELEMENT note định nghĩa phần tử note chứa 4 phần tử: to, from, heading, body
- ☐ !ELEMENT to định nghĩa phần tử to thuộc kiểu #PCDATA
- ☐ !ELEMENT from định nghĩa phần tử from thuộc kiểu #PCDATA
- ☐ !ELEMENT heading định nghĩa phần tử heading thuộc kiếu #PCDATA
- ☐ !ELEMENT body định nghĩa phần tử body thuộc kiểu #PCDATA

Khai báo DTD ngoại

- □ Nếu DTD được khai báo bên trong file bên ngoài, nó nằm trong định nghĩa DOCTYPE:
- <!DOCTYPE root-element SYSTEM "filename">

Ví du khai báo DTD ngoại

- <?xml version="1.0"?>
- <!DOCTYPE note SYSTEM "note.dtd">
- <note>
- <to>Tove</to>
- <from>Jani</from>
- <heading>Reminder</heading>
- <body>Don't forget me this weekend!</body>
- </note>

File note.dtd:

- <!ELEMENT note (to,from,heading,body)>
- <!ELEMENT to (#PCDATA)>
- <!ELEMENT from (#PCDATA)>
- <!ELEMENT heading (#PCDATA)>
- <!ELEMENT body (#PCDATA)>

Khai báo phần tử

- ☐ Phần tử <!ELEMENT> dùng để định nghĩa kiểu dữ liêu cho một phần tử của một tài liệu XML.
- □ Cú pháp:
 - <!ELEMENT element-name category>
 - hoăc
- <!ELEMENT element-name (element-content)>
- Phần tử với kiểu rỗng
 - <!ELEMENT element-name EMPTY>

 - Ví dụ: <!ELEMENT br EMPTY>
 - Ví du XML:
 -

10

Khai báo phần tử (tiếp)

- □ Phần tử có kiểu văn bản
 - <!ELEMENT element-name (#PCDATA)>
 - <!ELEMENT from (#PCDATA)>
- □ Phần tử với kiểu dữ liêu bất kỳ
 - <!ELEMENT element-name ANY>

 - <!ELEMENT note ANY>

Khai báo phần tử (tiếp)

- ☐ Đinh nghĩa phần tử có chứa một phần tử con
 - <!ELEMENT element-name (child-name)>
 - <!ELEMENT note (message)>
- □ Định nghĩa một phần tử chứa nhiều phần tử
 - Cách 1:
 - <!ELEMENT element-name (child1,child2,...)>
 - - <!ELEMENT note (to,from,heading,body)>

Khai báo phần tử (tiếp)

- <!ELEMENT note (to,from,heading,body)>
- <!ELEMENT to (#PCDATA)>
- <!ELEMENT from (#PCDATA)>
- <!ELEMENT heading (#PCDÁTA)>
- <!ELEMENT body (#PCDATA)>
- □ Chú ý, các phần tử con trong tài liệu có trình tự giống trình tự khai báo các phần tử con ở đây.
- ☐ Các phần tử con cũng có thể có các phần tử con mức dưới nữa.

Khai báo phần tử (tiếp)

- ☐ Cách 2: sử dụng kí tự đại diện
 - Phần tử có 1 hoặc nhiều phần tử con
 - <!ELEMENT element-name (child-name+)>
 - Ví du: <!ELEMENT note (message+)>
 - Phần tử có 0 hoặc nhiều phần tử con
 - <!ELEMENT element-name (child-name*)>

<!ELEMENT note (message*)>

Khai báo phần tử (tiếp)

- Phần tử có 0 hoặc 1 con
 - <!ELEMENT element-name (child-name?)>
 - <!ELEMENT note (message?)>
- Phần tử có một trong các phần tử con <!ELEMENT note (to,from,header,(message|body))>
 - phần tử note phải chứa phần tử to, from,
- header và {mesage hoặc body}
- Phần tử chứa phần tử con hoặc dữ liêu <!ELEMENT note (#PCDATA|to|from|header|message)*> phần tử note chứa 0 hoặc nhiều kiểu văn bản, phần tử to, from, header hoặc message.

15

Khai báo thuộc tính với <!ATTLIST>

- ☐ Phần tử <!ATTLIST> dùng để định nghĩa kiểu tư liệu của các thuộc tính cho một phần tử trong tài liệu XML
 - <!ATTLIST element-name attribute-name attribute-type default-value>
 - Ví du DTD:
 - <!ATTLIST payment type CDATA "check">
 - Ví dụ XML:
 - <payment type="check" />

16

Khai báo thuộc tính với <!ATTLIST>

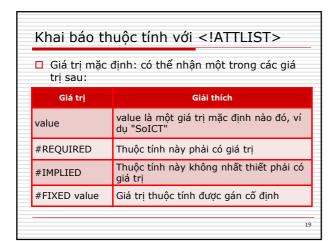
☐ Các kiểu thuộc tính có thể là:

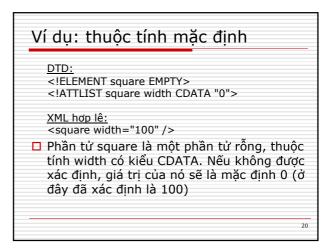
Kiểu	Mô tá
CDATA	Thuộc tính này chỉ có thể chứa kiểu dữ liệu kí tự
(en1 en2)	Thuộc tính chỉ nhận giá trị từ danh sách này
ID	Giá trị của thuộc tính này không được trùng nhau và bắt đầu bởi 1 chữ cái
IDREF	Giá trị của thuộc tính này phải là một trong các giá trị của thuộc tính ID của các phần tử khác
IDREFS	Giá trị của thuộc tính này phải là các giá trị của các thuộc tính có kiểu ID

Khai báo thuộc tính với <!ATTLIST>

Kiểu	Mô tả
NMTOKEN	Giá trị là một tên XML hợp lệ
NMTOKENS	Giá trị là một danh sách các tên XML hợp lệ
ENTITY	Giá trị là một thực thể
ENTITIES	Giá trị là một danh sách các thực thể
NOTATION	Giá trị là tên của một kí hiệu
xml:	Giá trị là một giá trị xml đã định sẵn

3





#REQUIRED

Cú pháp:

<!ATTLIST element-name attribute-name attributetype #REQUIRED>

Ví du:

DTD:

<!ATTLIST person number CDATA
#REQUIRED>

XML hơp lê:

<person number="5677" />
XML không hơp lê:

<person />

#IMPLIED

Cú pháp:
<!ATTLIST element-name attribute-name attributetype #IMPLIED>
Ví dụ:

DTD:
<!ATTLIST contact fax CDATA #IMPLIED>
XML hơp lê:
<contact fax="555-667788"/>
XML hơp lê:
<contact />

#FIXED

Cú pháp:

<!ATTLIST element-name attribute-name attribute-type #FIXED "value">

Ví du:

DTD:

<!ATTLIST sender company CDATA #FIXED "Microsoft">

XML hơp lê:

<sender company="Microsoft" />

XML không hợp lê:

<sender company="W3Schools" />

Thuộc tính nhận một trong các giá trị liệt kê

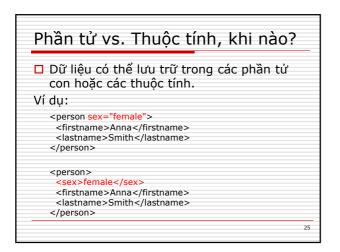
Cú pháp:
<!ATTLIST element-name attribute-name
(en1|en2|..) default-value>

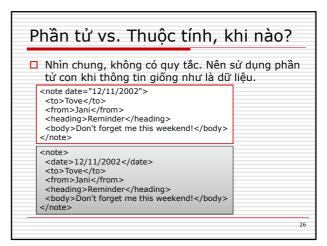
Ví du:

DTD:
<!ATTLIST payment type (check|cash) "cash">

Ví du XML:
<payment type="check" /> hoặc
<payment type="cash" />

124









Tránh sử dụng thuộc tính?

□ Ngoại lệ:

<messages>
<note id="p501">

<to>Trow
dẫm, bộ dịnh danh duy nhất, không phải là một cheading>Reminder</heading>
</note>
</note id="p502">
<to>Jon't forget me this weekend!</body>
cloud liệu note

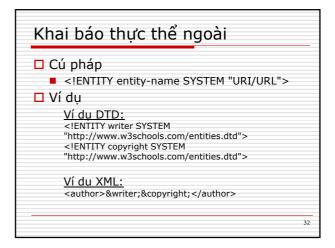
iệu note

iệu note

chody>I will not!</body>
</note>
</note>
</note>
</note>
</note id="p502">
<to>Jani</to>
</note>

Thực thể Các thực thể là các biến được sử dụng để định nghĩa shortcuts tới đoạn text chuẩn hoặc các kí tự đặc biệt. Các tham chiếu thực thể là các tham chiếu tới các thực thể Các thực thể có thể được khai báo bên trong hoặc bên ngoài



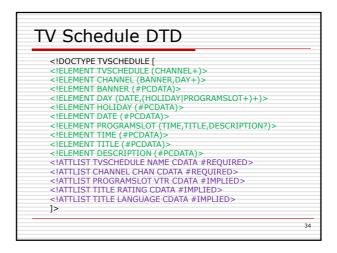


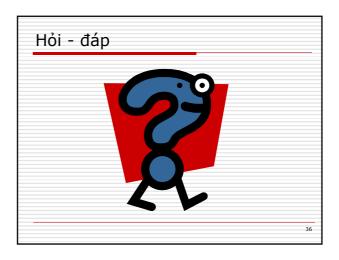
Một số ví dụ trên internet

TV Schedule DTD

By David Moisan. Copied from http://www.davidmoisan.org/
Newspaper Article DTD

Copied from http://www.vervet.com/





Lời hay ý đẹp "Thiếu tự tin là nguyên nhân của phần lớn mọi sự thất bại" Bowee

37