

XML

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

1

Nội dung

- Mở đầu
- Đơn vị xây dựng của tài liệu XML
- Các quy tắc
- Chỉ thị xử lý
- Dòng chú thích
- Không gian tên
- CDATA
- PCDATA
- Thực thể định nghĩa sẵn

2

Mở đầu

- eXtensible Markup Language
- để cấu trúc, lưu trữ và trong trao đổi dữ liệu giữa các ứng dụng và lưu trữ dữ liệu.
- Là ngôn ngữ xây dựng cấu trúc tài liệu văn bản, dựa theo chuẩn SGML (Standard Generalized Markup Language).
- Năm 1996, W3C (World Wide Web Consortium) thiết kế XML

3

Mở đầu

- XML vs HTML:
 - Đều có các tag (thẻ). Thẻ của HTML đã được định nghĩa trước và người dùng phải tuân thủ. Thẻ XML do người dùng định nghĩa
 - XML được thiết kế để chuyển tải và lưu trữ dữ liệu, tập trung vào "what data are". HTML được thiết kế để hiển thị dữ liệu, tập trung vào "how data look"
 - XML không phải là sự thay thế cho HTML.
- XML trở thành W3C Recommendation vào 10/02/1998
- XML đơn giản hóa việc chia sẻ và truyền tải dữ liệu, độc lập khi thay đổi platform.

4

Một ví dụ

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

5

Đơn vị xây dựng của tài liệu XML

- Phần tử
- Thuộc tính
- Thực thể
- PCDATA
- CDATA

6

Các quy tắc

- ❑ Phải có một phần tử gốc duy nhất chứa tất cả các phần tử khác trong tài liệu

```
<Catalog>
<Product productID="1">Chair</Product>
<Product productID="2">Desk</Product>
</Catalog>
```

- ❑ Mỗi tag mở phải có một tag đóng giống như nó.

7

Các quy tắc (tiếp)

- ❑ Các cặp ký tự của tag mở và tag đóng phải giống nhau hoàn toàn

```
<Order>
<OrderDate>2011-09-03</Orderdate>
<Customer>David Archuletta</Customer>
</Order>
```

- ❑ Mỗi phần tử con phải nằm trọn bên trong phần tử cha của nó
- ❑ Giá trị của thuộc tính phải được đặt trong cặp dấu nháy kép hoặc cặp dấu nháy đơn

8

Chỉ thị xử lý

- ❑ Dòng lệnh nằm ở đầu file

```
<?xml version="1.0" encoding="utf-8"
standalone="yes"?>
```

- ❑ Tài liệu XML có thể có hoặc không có phần này.

9

Dòng chú thích

- ❑ Đặt trong cặp <!-- và -->

10

Không gian tên (namespace)

- ❑ Có thể có nhiều tag trùng tên trong một tài liệu hay giữa các tài liệu có liên kết với nhau.

- ❑ Khai báo không gian tên

- Đưa thêm thuộc tính xmlns:prefix vào bên trong phần tử gốc. prefix là tên của không gian tên, mỗi không gian tên cần mang một định danh duy nhất. Một không gian tên có thể là một địa chỉ internet hoặc một địa chỉ nào đó (miễn là duy nhất!)

11

Ví dụ không gian tên hs

```
<?xml version="1.0"?>
<BookOrder
xmlns:hs="http://www.northwindtraders.com/customer">
<hs:Customer >
<hs>Title>Mr.</Title>
<hs:FirstName>Graeme</FirstName>
<hs:LastName>Malcolm</LastName>
</Customer>
</BookOrder>
```

12

Khai báo mặc định không gian tên

- Nếu tài liệu chỉ sử dụng một không gian tên thì khai báo mặc định, bỏ phần prefix

```
<?xml version="1.0"?>
<BookOrder xmlns="http://www.northwindtraders.com/customer">
  < Customer >
    < Title>Mr.</Title>
    < FirstName>Graeme</FirstName>
    < LastName>Malcolm</LastName>
  </ Customer>
</BookOrder>
```

13

Hai không gian tên mặc định

```
<?xml version="1.0"?>
<BookOrder>
  <OrderDate>2001-01-01</OrderDate>
  <Customer
    xmlns="http://www.northwindtraders.com/customer">
    <Title>Mr.</Title>
    <FirstName>Graeme</FirstName>
    <LastName>Malcolm</LastName>
  </Customer>
  <Book xmlns="http://www.northwindtraders.com/book">
    <Title>Treasure Island</Title>
    <Author>Robert Louis Stevenson</Author>
  </Book>
</BookOrder>
```

14

- Nếu có nhiều phần tử Customer và Book? Viết đi viết lại?
- Cách giải quyết

15

```
<?xml version="1.0"?>
<BookOrder xmlns="http://www.northwindtraders.com/order"
  xmlns:cus="http://www.northwindtraders.com/customer"
  xmlns:bok="http://www.northwindtraders.com/book">
  <OrderDate>2001-01-01</OrderDate>
  <cus:Customer >
    <cus:Title>Mr.</cus:Title>
    <cus:FirstName>Graeme</cus:FirstName>
    <cus:LastName>Malcolm</cus:LastName>
  </cus:Customer>
  <bok:Book >
    <bok:Title>Treasure Island</bok:Title>
    <bok:Author>Robert Louis Stevenson</bok:Author>
  </bok:Book>
</BookOrder>
```

16

CDATA

- CDATA nghĩa là character data
 - CDATA là đoạn text không được phân tích cú pháp bởi trình phân tích cú pháp.
 - Đoạn dữ liệu nằm giữa <![CDATA [và]]> sẽ được giữ nguyên khi đi qua trình phân tích → viết những đoạn script trong tài liệu
- ```
<script language="javascript">
 <![CDATA[
 function mag(){
 alert("This is CDATA!");
 }
]]>
</script>
```

17

## PCDATA

- PCDATA nghĩa là parsed character data
- PCDATA là đoạn text sẽ được phân tích cú pháp bởi trình phân tích cú pháp. Tag ở bên trong đoạn text sẽ được xử lý như là markup và các thực thể sẽ được expand.
- Dữ liệu kí tự không nên chứa &, <, >

18

## Thực thể định nghĩa sẵn

- XML sử dụng ký tự đặc biệt: <, >, ', ", &
- &apos; (nháy đơn ')
  - &amp; (dấu &)
  - &gt; (dấu >)
  - &lt; (dấu <)
  - &quot; (nháy kép ")

19

## Thực thể định nghĩa sẵn - Ví dụ:

```
<?xml version="1.0"?>
<LINK-TO>
< a href=" index.txt " > OPEN FILE INDEX.TXT
< /a >
</LINK-TO>

< a href=" index.txt " > OPEN FILE
INDEX.TXT < /a >
tương đương với
 OPEN FILE INDEX.TXT
```

20

## Hỏi - đáp



21

## Lời hay ý đẹp

"Sự khôn ngoan đầy quý giá có được là nhờ ở kinh nghiệm"

Roger Ascham

22