

Chương I: TỔNG QUAN XML

Giảng viên: Nguyễn Đức Cường - Khoa CNTT

Email: cuongnguyenduc@gmail.com

Website: <http://www.nguyenduccuong.com>

Chương 1 - Mục tiêu

- **Tổng quan XML.**
- **So sánh giữa HTML & XML**
- **Mô tả cấu trúc 1 file xml**
- **Các ứng dụng XML**
- **Soạn thảo và biên dịch XML**
- **Viết một tài liệu XML well-form**

Tổng quan XML

- XML- **Extensible Markup Language** - là ngôn ngữ tổng quát dùng để lưu trữ dữ liệu thông qua các thẻ có nghĩa.
- XML được định nghĩa bởi tổ chức mạng toàn cầu World Wide Web Consortium (W3C).
- Là ngôn ngữ đánh dấu độc lập với phần mềm, phần cứng và nền (cross-platform).
- Cho phép các máy tính truyền cấu trúc dữ liệu giữa hệ thống không đồng nhất.

XML được sử dụng ở đâu?

- Giảm tải server
- Nội dung web site
- Remote Procedure Calls
- E-Commerce

SO SÁNH GIỮA XML VÀ HTML [1]

- Cả HTML và XML đều dựa trên chuẩn ngôn ngữ định dạng tổng quát SGML

| XML | HTML |
|--|--------------------------------------|
| <i>Content</i> Markup Language | <i>Presenttation</i> Markup Language |
| Cho phép người dùng định nghĩa các phần tử | Các phần tử đã được định nghĩa sẵn |
| Kiểm tra tính hợp lệ | Hầu như không bắt lỗi |

SO SÁNH GIỮA XML VÀ HTML [2]

| XML | HTML |
|---|--------------------------------------|
| Cho phép trao đổi dữ liệu giữa các ứng dụng | Chỉ được thiết kế để trình diễn |
| Cần phải được định nghĩa một cách chặt chẽ | Không bắt buộc |
| Các phần tử phải được đóng | Các phần tử rỗng không cần phải đóng |

CẤU TRÚC CỦA XML

- Cấu trúc
- Ví dụ
- Ưu điểm của XML.

Cấu trúc tài liệu Xml

- Một tài liệu XML có 2 phần:
 - Cấu trúc logic: định nghĩa các phần tử, các thuộc tính, kiểu dữ liệu, vv...
 - Cấu trúc vật lý: chứa dữ liệu và các phần tử chứa dữ liệu như text, hình ảnh, media, vv...
 - ->Chú ý: tập tin Xml có phần mở rộng: .xml

Ví dụ: - greeting.xml

```
<?xml version="1.0" encoding="UTF-8"?>  
<?xml-stylesheet type="text/css" href="greeting.css"?>  
<document>  
  <greeting>  
    Hello  
  </greeting>  
  <message>  
    Welcome to my page  
  </message>  
</document>
```

Ví dụ - greating.css

```
greeting  
{  
  display: block; font-size: 30pt; color: blue  
}  
message  
{  
  display: block; font-size: 15pt; color: green  
}
```

Giải thích

`<?xml version="1.0" encoding="UTF-8"?>`

- Thẻ khai báo chỉ thị xử lý.
- Đặt ở ngay dòng đầu tiên.
- Trong đó:
 - `<document>` là phần tử gốc của tài liệu.
 - `<greeting>`, `<message>` là các phần tử được định nghĩa bên trong phần tử gốc, nó có thể chứa dữ liệu hoặc các phần tử khác.

Ưu điểm của XML

- Dễ dàng xử lý, chuyển tải và trao đổi dữ liệu.
- Mô tả dữ liệu và cách dữ liệu muốn thể hiện thông qua tên các thẻ.
- Tổ chức dữ liệu có cấu trúc và phân cấp.

CÁC ỨNG DỤNG THỰC TẾ CỦA XML

- CML - ngôn ngữ định dạng hoá học.
- MathML - ngôn ngữ định dạng toán học.
- SMIL (Synchronized Multimedia Integration Language).
- HTML+TIME
- XHTML: mở rộng và định nghĩa lại HTML.
- VML (Vector Markup Language) –
- WML – Wireless Markup Language.

Soạn thảo và biên dịch XML

- **Môi trường soạn thảo**
- **Biên dịch XML**

Môi trường soạn thảo XML

- XML Notepad

(www.microsoft.com/xml/notepad/intro.asp)

- XML Writer

(www.xmlwriter.net)

- XML Spy

(www.xmlspy.com)

- eNotePad

(www.edisys.com/products/enotepad)

Biên dịch và chạy tài liệu XML

- Bộ phân tích (XML Parser)
- Bộ kiểm tra (XML validator)
- Trình biên dịch

Bộ phân tích XML (XML PARSER)

- Tách dữ liệu XML thành từng mẫu nhỏ và khiến chúng có thể truy xuất được bởi mã lệnh của chương trình.
- Một số trình phân tích kiêm luôn cả chức năng kiểm tra khuôn dạng hoặc tính hợp lệ của tài liệu XML.
- **Các bộ phân tích thông dụng:**
 - XML for Java (XML4J), là thư viện phân tích tài liệu XML được phát triển bởi IBM AlphaWorks (www.alphaworks.ibm.com/tech/xml4j)
 - Microsoft XML Parser, được kèm theo sẵn trong trình duyệt Internet Explorer 5 trở lên.

Bộ kiểm tra (XML VALIDATOR)

- Dùng để kiểm tra xem tài liệu XML viết ra có đúng khuôn dạng và hợp lệ hay không?
- **Các bộ kiểm tra thường dùng:**
 - **XML Writer** là trình soạn thảo và kiểm tra tính hợp lệ của tài liệu.
 - **Tidy**: phân tích và kiểm tra tính hợp lệ của tài liệu. Địa chỉ: www.w3.org/people/raggett/tidy
 - Có thể kiểm tra tính hợp lệ của tài liệu ngay trên Internet qua địa chỉ: www.stg.brown.edu/service/xmlvalid

Trình duyệt

- Internet Explorer
- Mozilla FireFox
- Netscape navigator
- ...

XÂY DỰNG TÀI LIỆU XML

- Ví dụ: tạo một tài liệu XML để lưu thông tin về đơn đặt hàng của các khách hàng cụ thể.

- Chỉ thị xử lý:

```
<?xml version="1.0" standalone="yes"?>
```

`standalone="yes"`: cho biết tài liệu là một thực thể duy nhất không tham chiếu đến các thực thể khác. Đây cũng là giá trị mặc định.

- Tạo phần tử gốc:

```
<?xml version="1.0" standalone="yes"?>
```

```
<document>
```

```
...
```

```
</document>
```

VÍ DỤ (tt)

- Tạo các phần tử con

```
<?xml version="1.0" standalone="yes"?>
```

```
<document>
```

```
  <customer>
```

```
    ...
```

```
  </customer>
```

```
  <customer>
```

```
    ...
```

```
  </customer>
```

```
</document>
```

VÍ DỤ (tt)

- Đối với mỗi khách hàng ta thêm thông tin về tên bằng thẻ `<name>` cùng với thông tin chi tiết `<first_name>` và `<last_name>`

```
<?xml version="1.0" standalone="yes"?>
```

```
<document>
```

```
  <customer>
```

```
    <last_name>Smith</last_name>
```

```
    <first_name>Sam</first_name>
```

```
  </customer>
```

```
  <customer>
```

```
    ...
```

```
  </customer>
```

```
</document>
```

VÍ DỤ (tt)

- Thêm thông tin về đơn đặt hàng

```
<customer>
```

```
  <last_name>Smith</last_name>
```

```
  <first_name>Sam</first_name>
```

```
  <date>October 15, 2001</date>
```

```
  <orders>
```

```
    ...
```

```
  </orders>
```

```
</customer>
```

VÍ DỤ (tt)

```
<customer>
  <last_name>Smith</last_name>
  <first_name>Sam</first_name>
  <date>October 15, 2001</date>
  <orders>
    <item>
      <product>Tomatoes</product>
      <number>8</number>
      <price>$1.25</price>
    </item>
  </orders>
</customer>
```

XML WELL - FORM

- XML phân biệt chữ hoa và chữ thường
- Các khai báo XML phải được đặt ngay dòng đầu tiên của tài liệu
- Tài liệu chỉ có 1 phần tử gốc, các phần tử khác nếu có phải là con của phần tử gốc
- Mọi phần tử XML khác rỗng phải bao gồm đầy đủ thẻ đóng
- Đóng phần tử rỗng với chuỗi đóng />
- Mọi phần tử trong tài liệu khác phần tử gốc đều phải nằm giữa cặp thẻ gốc
- Các giá trị của thuộc tính phải được đặt trong dấu nháy đôi (“”)

Một số khái niệm khác

- Tập ký tự và mã hóa
- Không gian tên (Namespace)

TẬP KÝ TỰ VÀ MÃ HOÁ

- Tập ký tự mặc định của XML là UTF-8 (8 bits). Đây là bộ mã thu gọn của UTF-16 (16 bits), nó chính là bộ mã ASCII nằm trong phần đầu của bộ mã Unicode.

KHÔNG GIAN TÊN (NAMESPACES)

- Khi kết hợp các tài liệu XML với nhau có thể xảy ra trường hợp trùng tên thẻ, vì vậy cần định nghĩa không gian tên để phân biệt chúng.

TẠO KHÔNG GIAN TÊN (tt)

■ Ví dụ:

Đây là một thư viện sách có phần tử gốc là <library>

```
<library>
```

```
  <book>
```

```
    <title>XML</title>
```

```
  </book>
```

```
</library>
```

VÍ DỤ (tt)

Không gian tên dcs được tạo ra cho thư viện như sau:

```
<library
```

```
  xmlns:dcs="http://www.incoll4.edu.vn/lib">
```

```
  <book>
```

```
    <title>XML</title>
```

```
  </book>
```

```
</library>
```

VÍ DỤ (tt)

- Áp dụng không gian tên dcs cho tất cả các thẻ nào trong tài liệu mà ta muốn chúng duy nhất và độc quyền về dcs

```
<dcs:library
```

```
  xmlns:dcs="http://www.incoll4.edu.vn/lib">
```

```
    <dcs:book>
```

```
      <dcs:title>XML</dcs:title>
```

```
    </dcs:book>
```

```
</dcs:library>
```

VÍ DỤ (tt)

- Nếu ta cần tham khảo đến các tựa sách của Amazon, tài liệu này chắc chắn sẽ có phần tử mang tên <book>. Để an toàn ta sẽ tạo một không gian tên cho Amazon như sau:

```
<dcs:library
```

```
  xmlns:dcs="http://www.incoll4.edu.vn/lib"
```

```
  xmlns:amazon="http://www.amazon.com/lib"
```

```
  <dcs:book>
```

```
    <dcs:title>XML</dcs:title>
```

```
  </dcs:book>
```

```
</dcs:library>
```

VÍ DỤ (tt)

- **Bây giờ bất kỳ tựa sách nào của Amazon đều có thể đưa vào với cùng định dạng và cấu trúc của thư viện dcs**

```
<dcs:library
```

```
  xmlns:dcs="http://www.incoll4.edu.vn/lib"
```

```
  xmlns:amazon="http://www.amazon.com/lib"
```

```
  <dcs:book>
```

```
    <dcs:title>XML</title>
```

```
  </dcs:book>
```

```
  <amazon:book>
```

```
    <amazon:title>XML</amazon:title>
```

```
  </amazon:book>
```

```
</dcs:library>
```

CDATA

- Thừa kế từ SGML - Character DATA
- XML parser sẽ bỏ qua việc phân tích văn bản.

`<tagname><![CDATA[text]]></tagname>`