**Cấu trúc của HTML**

**HTML** (viết tắt của từ **HyperText Markup Language**, hay là "Ngôn ngữ Đánh dấu Siêu văn bản") là một [ngôn ngữ đánh dấu](https://vi.wikipedia.org/wiki/Ng%C3%B4n_ng%E1%BB%AF_%C4%91%C3%A1nh_d%E1%BA%A5u) được thiết kế ra để tạo nên các [trang web](https://vi.wikipedia.org/wiki/Website) trên [World Wide Web](https://vi.wikipedia.org/wiki/World_Wide_Web).

Được trợ giúp bởi các công nghệ như [CSS](https://vi.wikipedia.org/wiki/CSS) và các ngôn ngữ kịch bản giống như [JavaScript](https://vi.wikipedia.org/wiki/JavaScript).

Ứng dụng :

-Phần mềm ứng dụng trên nền tảng Website (Web app), Web mail , Web game , …

**Cấu trúc** cơ bản trang của HTML có dạng:

<!DOCTYPE html> phần khai báo kiểu dữ liệu hiển thị là HTML để trình duyệt biết

<html> cặp thẻ nằm ngoài cùng bao hết nội dung của trang web lại

<head> phần khai báo tt trang web

<title> Tiêu đề trang </title> khai báo tiêu đề

<meta charset="utf-8"> Ngôn ngữ hỗ trợ được khai báo tại đây

*// Với utf - 8 là ngôn ngữ tiếng Việt. UTF-8 ở đây là viết tắt của 8-bit Unicode Transformation Format có nghĩa là định dạng chuyển đổi Unicode 8-bit. Nó được sử dụng như một bộ mã hóa ký tự với chiều rộng biến thiên dành cho Unicode. //*

</head>

<body> trong này chứa những đoạn mã html để hiển thị trên trang web

<h1> Tiêu đề </h1> tiêu đề lớn nhất

(h1-h6: heading = tiêu đề, phông chữ giảm dần <h1>heading</h1>)

<p> Đoạn văn </p>

<div> Chứa các khối , trong khối có nhiều câu lệnh </div>

<h2> Tiêu đề </h2>

</body>

</html> đóng thẻ HTML

- Nội dung của một tập tin HTML phải được bắt đầu bằng thẻ <html> và kết thúc bởi thẻ </html>.

- Bên trong cặp thẻ <html></html> được chia làm hai thành phần chính:

Phần head được bắt đầu bằng thẻ <head> và kết thúc bởi thẻ </head>, nó dùng để chứa tiêu đề và các thẻ khai báo thông tin cho trang web.

Phần body được bắt đầu bằng thẻ <body> và kết thúc bởi thẻ </body>, nó dùng để chứa những nội dung mà chúng ta muốn hiển thị trên trang web.

**Thư viện** hỗ trợ HTML : BeautifulSoup

**Cấu trúc file XML**

XML là từ viết tắt của từ Extensible Markup Language là ngôn ngữ đánh dấu mở rộng. XML có chức năng truyền dữ liệu và mô tả nhiều loại dữ liệu khác nhau. Tác dụng chính của XML là đơn giản hóa việc chia sẻ dữ liệu giữa các nền tảng và các hệ thống được kết nối thông qua mạng Internet.

Cấu trúc XML được dựa trên các NODE lồng vào nhau. Mỗi NODE có đủ 1 thẻ đóng và 1 thẻ mở.

Ví dụ cấu trúc :

<NodeName>Nội dung</NodeName>

Một cách ứng dụng nữa là thêm thuộc tính cho thẻ như sau:

<NodeName> Attribute=””Nội dung</NodeName>

Cấu trúc file XML gồm có 5 thành phần chính:

Header file:

Nó được biết đến như thẻ đầu tiên trong file XML. Header file được dùng để thông báo version (thường là 1.0) và còn có tác dụng lưu trữ thông tin về việc mã hóa ký tự. Giá trị encoding của Header file sẽ thuộc một trong những dạng sau: UTF-8, UTF-16, ISO-10646-UCS-2, ISO-10646-UCS-4, ISO-2022-JP. Còn về encoding là gì thì nó là một dạng mã hóa ký tự.

Ví dụ về thẻ Header file:

<?xml version=”1.0″ encoding=”utf-8″ ?>.

Root node: Là phần tử bên ngoài cùng hay còn được gọi là phần tử gốc. Nó thường là tên gọi hoặc đại diện cho thông tin lưu trữ bên trong XML.

Element: Cũng là một dạng tên gọi nhưng sẽ dùng để chỉ các tệp con nằm trong Root node.

Attribute: Là thuộc tính của thẻ, được dùng để phân biệt đặc tính giữa 2 element trùng tên.

Employee: Mỗi thẻ Employee đại diện cho một nhân viên. Nhiều thẻ Employee sẽ tạo nên một tổ hợp Employees. Trong thẻ Employee còn có những thẻ con để chỉ các thông tin nhỏ lẻ của nhân viên như EmpID, Name, Phone…

Những cách ứng dụng phổ biến của XML

XML trở nên phổ biến như vậy là nhờ rất nhiều ứng dụng ưu việt của nó. Để học XML, bạn có thể tham khảo những cách sử dụng web.xml như sau:

***Xuất bản web***

XML được sử dụng như một công cụ thiết kế web. Các dữ liệu được tạo ra và lưu trữ trong XML. Người dùng có thể sử dụng chúng trên nhiều máy khác nhau. Để chuyển đổi XML thành những bố cục khác, ví dụ như HTML cho các trang web, bạn chỉ cần sử dụng các bộ xử lý chuyển đổi đơn giản. Nhà phát triển web sẽ sử dụng XML như một công cụ hỗ trợ cho hoạt động chỉnh sửa và thao tác nội dung cho các trang web.

***Tác vụ web***

XML giúp tối ưu hóa khả năng tìm kiếm và tự động hóa các tác động trên web. Nó cho phép người dùng tìm kiếm nhanh chóng một thông tin nào đó với các kết quả rất chuẩn xác. Ví dụ, nếu bạn tìm kiếm cụm từ Black Widow (một loài nhện) qua HTML, kết quả trả về sẽ có cả Black – màu đen và Widow – góa phụ. Nhưng nếu bạn tìm kiếm qua XML, kết quả được thu gọn lại, thông tin cô đọng và chính xác hơn.

Thư viện hỗ trợ XML : BeautifulSoup

## **Ứng dụng file XML trong thực tiễn**

XML thường được dùng làm cơ sở cho các định dạng tài liệu khác nhau. Một số định dạng bạn có thể biết như:

* ATOM và RSS: mô tả cách ứng dụng đọc xử lý nguồn cấp web.
* Microsoft Office 2007 và các phiên bản tiếp theo đều sử dụng XML làm cơ sở cấu trúc tài liệu. Một số định dạng tài liệu bạn có thể thấy như .docx, chữ “X” được sử dụng xuất phát từ file XML. Bên cạnh đó, bạn cũng có thể gặp định dạng này trong excel và powerpoint.
* Microsoft .NET cũng sử dụng **XML** cho các file có cấu hình nhỏ của nó.

Tất cả các loại ứng dụng đều có thể hưởng lợi từ XML vì nó cung cấp một phương pháp truy cập thông tin được sắp xếp hợp lý. Quá trình đơn giản này cho phép cả ứng dụng và thiết bị sử dụng, lưu trữ, truyền và hiển thị dữ liệu. Ví dụ, ở nơi làm việc, các kiến ​​trúc sư và lập trình viên dữ liệu sử dụng XML hàng ngày.