**Мова XML та аналіз XML-файлів**

Мова XML використовується для створення структури даних, щоб потім передавати або зберігати їх. Вона популярна у всіх сферах програмування, оскільки відрізняється простотою сприйняття та універсальністю прочитання різними додатками.

**Опис мови XML**

XML (розширювана мова розмітки) - це мова програмування, яка складається з оголошень у вигляді інформації та визначальних тегів. З її допомогою зручно зберігати та передавати будь-які дані.

Мова не залежить від операційної системи та середовища обробки. XML служить для уявлення деяких даних як структури, які ви можете самі розробити чи підлаштувати під програму чи сервіс.

Саме тому цю мову називають розширюваною, і в цьому її головна перевага, за яку її так цінують.

Плюси мови XML:

* легкість читання, подання у простій формі;
* стандартний вид кодування;
* можливість створення різних структур (списків, схем, дерев);
* можливість відновити дані, збережені в XML;
* можливість обміну даними між будь-якими платформами;
* популярність у різних сферах програмування.

Мінуси мови XML:

* надмірний синтаксис, велика кількість сутностей та тегів;
* один об'єкт може бути представлений у різних описах;
* відсутні стандартні вказівки типу об'єкта.

**Структура XML**

Мова програмування XML може використовуватися для групування будь-яких даних для створення ієрархії або розмітки.

Структура XML представлена ​​простим самописним синтаксисом:

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

<marvel>

<!-- this is a good man -->

<hero id="positive\_character">

<nickname>Captain America</nickname>

<realname>Steven Rogers</realname>

<abilities>Superhuman strength</abilities>

</hero>

<!-- this is a bad man -->

<hero id="negative\_character">

<nickname>Red Skull</nickname>

<realname>Johann Schmidt</realname>

<abilities>Superhuman strength</abilities>

</hero>

</marvel>

Тепер розберемо цей приклад докладніше.

Перший рядок документа – це XML декларація. Тут визначається версія XML (version="1.0") та тип кодування документа (encoding="UTF-8").

Далі описується кореневий елемент документа. Кореневий елемент у документі може бути лише один, і він міститиме всі ваші дані. Оскільки ми можемо самі назвати тегами, ми використовували <marvel>...</marvel>. Наш документ міститиме список героїв <hero>...</hero>.

Зауважте, що в документі можна писати коментарі, але це не обов'язкова вимога.

Далі всередині тегу <hero>...</hero> ми описуємо кожного героя. У нас є тег <nickname>...</nickname>, у який ми записали псевдонім героя, <realname>...</realname> — у який ми записали реальне ім'я героя, <abilities>...</abilities> - в який ми записали суперздатності героя.

Також у тега hero є атрибут (id="positive\_character"). Атрибути надають додаткову інформацію щодо елемента. Ця інформація може бути важливою для додатків, які маніпулюватимуть цим елементом. Значення атрибута завжди повинне полягати в лапки. Назву атрибута ви можете придумати самі, оскільки мова - розширюється.

Синтаксис XML виглядає просто, але не приймає помилок. Наприклад, якщо ви пропишите значення атрибута без лапок, це викликає синтаксичну помилку, тому обов'язково валідуйте свій файл.

**Для чого використовується мова XML**

XML може застосовуватись у будь-якій сфері програмування. Наприклад, ієрархії з XML-даними можуть використовуватися в:

* XHTML – для відображення сторінок в інтернеті;
* SVG - для опису картинок у векторному форматі;
* RDF – для опису ресурсів каталогів;
* WSDL – для звернення до віддалених програм та веб-сервісів;
* OWL - для опису структури каталогів;
* XAML – для опису інтерфейсу програм.

Синтаксис XML також можна використовувати для представлення математичних символів, медичних формул, синтезаторів мови і т.д.

**Додатки XML**

* XPath - неймовірно гнучкий, потужний і простий інструмент для навігації за документами XML. Використовується для пошуку запитів до елементів;
* XQuery — мова запитів, розроблена для обробки даних у форматі XML;
* XSLT — використовується для трансформації документів XML в інші формати (наприклад, для трансформації XML в HTML).

**XML і HTML**

XML не є заміною HTML. Ці дві мови доповнюють одна одну.

HTML використовується для відображення інформації, а XML - для зберігання та транспортування інформації.

У XML ви винаходите свої власні теги, оскільки в мові немає певних тегів. У HTML всі теги, що використовуються, визначаються стандартами HTML.

У документах XML можуть міститися будь-які необов'язкові описи, щоб інші додатки могли виконати перевірку його структури.

Документ XML можна перетворити на формат HTML. При передачі даних із сервера до браузера цей процес виконується автоматично. Щоб вручну конвертувати XML у HTML, можна використовувати онлайн-інструменти.

Також можна виконати зворотне перетворення HTML на XML.

**Як відобразити код XML**

Відкрити документ у цьому форматі можна будь-яким із зазначених способів:

* Блокнот Windows;
* Браузер;
* Microsoft Excel;
* Microsoft Word;
* Спеціалізовані редактори XML, наприклад, XML Editor.

**Висновок**

Загалом, розуміння мови XML та навички аналізу XML-файлів можуть бути корисними для багатьох сфер, включаючи веб-розробку, бази даних, аналітику даних та багато іншого. Ця мова використовується у всіх галузях програмування і дуже популярна як метод перетворення об'ємної інформації у форму ієрархії для її зручного зберігання.