**Расширяемый язык разметки XML**

1. Определение
2. Структура и особенности
3. Правила создания XML документов
4. Комментарии и атрибуты
5. Специальные символы и директива анализатора
6. CDATA

**1. Определение XML**

XML (eXtensible Markup Language) — расширяемый язык разметки, предназначенный для хранения и передачи структурированных данных в текстовом формате. Он не содержит операторов, не выполняет вычислений и используется как метаязык — то есть как основа для создания собственных языков описания данных.

**2. Структура и особенности**

**Структура XML-документа:**

* **Пролог** — строка вида <?xml version="1.0" encoding="UTF-8"?>, указывает версию и кодировку (необязательный, но рекомендуется).
* **Корневой элемент** — единственный главный тег, внутри которого находятся все остальные.
* **Элементы** — пары открывающего и закрывающего тегов, могут быть вложенными.
* **Атрибуты** — дополнительные параметры внутри открывающего тега, например: <book id="1">.
* **Текстовое содержимое** — данные между тегами.
* **Правильная вложенность** — теги не должны перекрываться, структура должна быть строго иерархической.
* **Чувствительность к регистру** — <Book> и <book> считаются разными элементами.

**Особенности XML:**

* **Гибкость** — можно создавать собственные теги и структуры.
* **Расширяемость** — подходит для описания любых типов данных.
* **Читаемость** — понятен как человеку, так и машине.
* **Стандартизация** — поддерживается большинством языков программирования и платформ.
* **Валидация** — можно проверять структуру через DTD или XSD.
* **Формат хранения** — используется для конфигурационных файлов, обмена данными, веб-сервисов и других задач.

**3. Правила создания XML-документов**

* Документ должен начинаться с объявления XML (пролог).
* Каждый открывающий тег должен иметь соответствующий закрывающий.
* Регистр символов имеет значение.
* Значения атрибутов должны быть заключены в кавычки.
* Вложенность тегов должна быть строго соблюдена.
* Все символы между тегами считаются данными — пробелы, табуляции и переводы строк сохраняются и не игнорируются, как это бывает в HTML.

**4. Комментарии и атрибуты**

**Атрибуты** — это пары вида имя="значение", задаются внутри открывающего тега и уточняют характеристики элемента. Пример: <book title="XML Basics">.

**Комментарии** — используются для пояснений внутри кода, не обрабатываются парсером: <!-- Это комментарий -->

**5. Специальные символы и директива анализатора**

**Специальные символы** — чтобы избежать конфликтов с синтаксисом XML, используются символьные или числовые представления:

* < → &lt;
* > → &gt;
* " → &quot;
* ' → &apos;
* & → &amp;
* Также возможны числовые формы: &#36; (десятичная) или &#x1A; (шестнадцатеричная).

**Директивы анализатора** — специальные инструкции для XML-парсера, например:

* Определение типа документа (<!DOCTYPE ...>)
* Пространства имён (xmlns:prefix="URI")

**6. CDATA**

**CDATA-секция** — используется для включения текста, который не должен интерпретироваться как разметка. Синтаксис:

<![CDATA[

<sender>John Smith</sender>

]]>

Всё, что находится внутри <![CDATA[ ... ]]>, воспринимается как обычный текст, даже если содержит символы <, >, & и другие. Это удобно, если нужно вставить фрагмент кода или текст с потенциально конфликтующими символами.