

Введение в XML

№ урока: 1 **Курс:** Основы Java EE

Средства обучения: JDK, IntelliJ Idea

Обзор, цель и назначение урока

Введение в XML. Валидация.

Изучив материал данного занятия, учащийся сможет:

- Научиться создавать свой XML-файл.
- Научиться создавать схемы для XML-файла.
- Разберется, зачем нужно пространство имен.

Содержание урока

1. Рассмотрение XML.
2. Рассмотрения пространства имен XML
3. Рассмотрение синтаксиса XML.
4. Рассмотрение DTD-схемы XML.
5. Рассмотрение XSD-схемы XML

Резюме

- **XML** — расширяемый язык разметки. Рекомендован Консорциумом Всемирной паутины (W3C). Спецификация XML описывает XML-документы и частично описывает поведение XML-процессоров (программ, читающих XML-документы и обеспечивающих доступ к их содержимому). XML разрабатывался как язык с простым формальным синтаксисом, удобный для **создания** и обработки документов программами и одновременно удобный для чтения и создания документов человеком, с подчёркиванием нацеленности на использование в Интернете. Язык называется расширяемым, поскольку он не фиксирует разметку, используемую в документах: разработчик волен создать разметку в соответствии с потребностями к конкретной области, будучи ограниченным лишь синтаксическими правилами языка. Расширение XML — это конкретная грамматика, созданная на базе XML и представленная словарём тегов и их атрибутов, а также набором правил, определяющих какие атрибуты и элементы могут входить в состав других элементов. Если бы HTML появился позднее, чем XML, то он, возможно, также бы являлся его расширением. Сочетание простого формального синтаксиса, удобства для человека, расширяемости, а также базирование на кодировках Юникод для представления содержания документов привело к широкому использованию как собственно XML, так и множества производных специализированных языков на базе XML в самых разнообразных программных средствах.
- **Пространство имён** в XML — это стандарт, описывающий именованную группу имён элементов и атрибутов, служащую для обеспечения их уникальности в XML-документе. В декабре 2009 года третья редакция стандарта получила статус рекомендации. Все имена элементов в пределах пространства имён должны быть уникальны. XML-документ может содержать имена элементов и атрибутов из нескольких словарей XML. В каждом словаре задано своё пространство имён — так разрешается проблема неоднозначности имён элементов и атрибутов. Возьмём, например, простой XML, который содержит ссылки на покупателя и на заказанный продукт. И элементы покупателя, и элементы продукта могут содержать дочерний элемент «ID_number». Ссылки на элемент ID_number будут, таким образом, неоднозначны, у нас будут два одинаковых имени элемента, несущих разную смысловую нагрузку и так будет, пока мы не введём пространства имён для их различения.

- XML Schema** — язык описания структуры XML-документа. Спецификация XML Schema является рекомендацией W3C. Как большинство языков описания XML, XML Schema была задумана для определения правил, которым должен подчиняться документ. Но, в отличие от других языков, XML Schema была разработана так, чтобы её можно было использовать в создании программного обеспечения для обработки документов XML.
 После проверки документа на соответствие XML Schema читающая программа может создать модель данных документа, которая включает:
 - словарь (названия элементов и атрибутов);
 - модель содержания (отношения между элементами и атрибутами и их структура);
 - типы данных.
 Каждый элемент в этой модели ассоциируется с определённым типом данных, позволяя строить в памяти объект, соответствующий структуре XML-документа. Языкам объектно-ориентированного программирования гораздо легче иметь дело с таким объектом, чем с текстовым файлом.
 Другим удобством XML Schema является то, что один словарь может ссылаться на другой, и, таким образом, разработчик может использовать уже существующие словари и легче устанавливать и распространять стандарты XML структуры для определённых задач (например, словарь протокола SOAP).
- DTD** — включает в себя два понятия:
 Термин, который используется для описания схемы документа или его части *языком схем DTD*.
 Язык схем DTD (DTD schema language) — компьютерный язык, который используется для записи фактических синтаксических правил метаязыков разметки текста SGML и XML. С момента его внедрения другие языки схем для спецификаций, такие как XML Schema и RELAX NG, выпускаются с дополнительной функциональностью.
 Из-за определённых отличий между XML и SGML, применение DTD также имеет некоторые особенности в зависимости от целевого документа.
 Сейчас идёт отказ от использования DTD в XML-технологии по ряду причин:
 - Используется отличный от XML синтаксис.
 - Отсутствует типизация узлов.
 - Отсутствует поддержка пространств имён.
 На смену DTD пришёл стандарт консорциума W3C XML Schema.
- RELAX NG (REgular Language for XML Next Generation)** — один из языков описания структуры XML-документа. Являясь сам по себе XML-документом, схема в этом формате может быть записана с использованием альтернативного, более компактного синтаксиса. В сравнении с другими языками схем, RELAX NG относительно прост. RELAX NG был разработан в OASIS и впервые опубликован в 2003. Файлы, содержащие схемы RELAX NG, обычно имеют расширение ".rng", а в компактном синтаксисе — ".rnc".

Закрепление материала

- Что такое XML?
- Зачем нужны схемы в XML?
- Что такое пространство имен и зачем оно нужно?

Дополнительное задание

Задание

Создать XML файл согласно заданию на стр.449 выбрать вариант на выбор.

https://careers.epam.by/content/dam/epam/by/book_epam_by/JAVA_Methods_Programming_v2.march2015.pdf

Сгенерировать с помощью IntelliJ IDEA XSD-схему. Используя стандартные средства для работы с файлами вывести в консоль содержимое XML-файла.

Самостоятельная деятельность учащегося

Задание 1

Выучите основные понятия, рассмотренные на уроке.

Задание 2

Используя XML из дополнительного задания переписать XSD так, чтобы использовались простые и комплексные типы, перечисления, предельные значения.

Задание 3

Используя работу с файлами вывести не все содержимое XML-документа, а только то, что находится внутри тега. (Подсказка: используйте регулярные выражения).

Задание 4

Используя поисковые механизмы <https://www.w3schools.com/> найти материал по теме и закрепить полученные знания.

Рекомендуемые ресурсы

XML

<https://www.w3schools.com/xml/>

Namespaces

https://www.w3schools.com/xml/xml_namespaces.asp

DTD

https://www.w3schools.com/xml/xml_dtd.asp

XML Scheme

https://www.w3schools.com/xml/xml_schema.asp