

XML, Сериализация в XML

Java Collections
3 уровень, 6 лекция

ОТКРЫТА

— Как делишки?

— Отлично. Не жалуюсь. Сегодня Билабо рассказал про JavaScript. Не все, конечно, но тоже немало. Я конечно еще на JS ничего не писал, но думаю, что это несложно.

А Элли рассказала про сериализацию в JSON. И ты рассказывал о Jackson framework и как настраивать «полиморфную десериализацию» с помощью аннотаций.

— Нифига себе, ты умный теперь, Амиго! Серьезный мужик!

— А то!

— Ладно. Давай к работе. Сегодня **новая и интересная тема – XML**.



XML – это стандарт представления данных, которые легко могут быть прочитаны человеком, и еще легче – программой.
Пример xml-файла:

XML	
1	<data>
2	<owner first="Nikolay" last="Ivanovich">
3	<address>Moscow</address>
4	</owner>
5	<cat name="Murka" age="15"/>
6	</data>

Основа XML – это теги. Тег – это слово в треугольных скобках (знаки меньше и больше). Теги бывают открывающие и закрывающие. Одному открывающему всего соответствует один закрывающий тег. Открывающие теги могут иметь атрибуты.

В тег можно складывать другие теги, получая, таким образом, дерево элементов. Тег верхнего уровня называют корнем (root): у него есть дочерние теги, у них свои дочерние теги.

Примеры:

Тег	Описание
<data>	Открывающий тег «data»
</data>	Закрывающий тег «data»

НАЧАТЬ ОБУЧЕНИЕ

<div><data></div> <div><owner></div> <div><cat name="Murka"/></div> <div></owner></div> <div></data></div>	Вложенные теги.
<div><cat name="Murka" age="15" /></div>	Автозакрывающийся тег. Такому тегу не нужен закрывающий. Не может содержать дочерние теги.
<div><info></div> <div>Тут может быть любая информация</div> <div></info></div>	Тег может содержать текстовые данные
<div><info></div> <div>Тут может</div> <div><data xxx="yyy"></div> <div></data></div> <div>быть любая</div> <div><data 2xxx="yyy"/></div> <div>информация</div> <div></info></div>	Тег может содержать текстовые данные и другие теги вперемешку.

— Выглядит несложно. А какие теги бывают?

— Любые. Никаких зарезервированных тегов нет. XML – это язык, для описания любых данных. Люди сами придумывают теги для своих нужд и договариваются, как их использовать.

Фактически, XML – это способ записать данные в виде дерева элементов, понятный компьютеру.

— Вроде ясно. Кстати, у меня вопрос.

Вот JSON используется для передачи данных с браузера на сервер, а где используется XML?

— Да там же, где и JSON: для хранения и передачи данных.

Ладно, продолжим.

Представь, что есть один общий XML-файл, который хранит данные программы, которую пишут двадцать человек. Тогда каждый из них придумывает теги для себя и очень скоро они начинают мешать друг другу.

Чтобы гарантировать уникальность тегов, им были придуманы префиксы. Выглядит это так:

Теги	Описание
<div><animal:cat></div>	Тег cat с префиксом animal
<div><animal:cat></div> <div></animal:cat></div> <div><zoo:cat></div> <div></zoo:cat></div>	Два тега cat с разными префиксами.
<div><animal:cat zoo:name="MX"></div>	Тег cat с префиксом animal , атрибут name с префиксом zoo .

Префиксы еще называют **namespace – пространство имен**. Тогда последнее предложение в таблице будет звучать так «Тег cat из пространства имен animal с атрибутом name из пространства имен zoo».

НАЧАТЬ ОБУЧЕНИЕ

Кстати, помнишь в Java у класса есть короткое имя, а есть длинное уникальное имя, в которое входит название пакета, его еще указывают при импорте?

— Ага.

— Так вот, у префиксов тоже есть уникальное длинное имя и его тоже указывают при импорте:

Пример
<pre><data xmlns:soap="http://cxf.apache.org/bindings/soap"> <soap:item> <soap:info/> </soap:item> </data></pre>

«xmlns:soap» значит «XML—namespace SOAP»

Более того, если есть теги без префикса, можно задать и их уникальное имя:

Пример
<pre><data xmlns="http://www.springframework.org/schema/beans" xmlns:soap="http://cxf.apache.org/bindings/soap" xmlns:task="http://www.springframework.org/schema/task"> <soap:item> <soap:info/> <task:info/> </soap:item> </data></pre>

«xmlns=...» задает namespace для пустого префикса – т.е. для тегов без префикса, таких как data.

В документе может быть сколько угодно namespaces, но у каждого должно быть свое уникальное имя.

— Ясно. А почему такие странные «уникальные имена» у этих namespaces?

— Обычно там указывают URL, по которому находится документ, который описывает этот namespace и/или его XML-теги.

— Не так уж и мало на меня сегодня вылили информации. Что еще?

— Еще немного осталось.

Во-первых, у XML есть заголовок – специальная строка, которая описывает версию этого XML и кодировку файла.

Выглядит она обычно так:

Пример
<pre><?xml version="1.0" encoding="UTF-8"?> <data xmlns:soap="http://cxf.apache.org/bindings/soap"> <soap:item> <soap:info/> </soap:item> </data></pre>

Еще в XML можно вставлять комментарии. Для начала комментария используют «<!--», а для конца «—>».

Пример
<pre>1 <?xml version="1.0" encoding="UTF-8"?> 2 <data xmlns:soap="http://cxf.apache.org/bindings/soap"></pre>

```
4      <!-- <soap:info/> -->
5      </soap:item>
6      <!-- это тоже комментарий -->
7      </data>
```

— Пока понятно.

— Из-за того, что в XML используются символы « < > “ & », их нельзя использовать в других местах. Для их описания используют так называемые «эскейп последовательности» — набор символов для представления других символов/символа. Вот список некоторых из них:

Escape-последовательность	Символ, который она заменяет
&	&
"	"
<	<
>	>
'	'

А вот пример кода, который хранится в XML:

Java-код	Он же, в XML
<div><div>1</div><div>if (a<b)</div></div> <div><div>2</div><div>System.out.println("a is minimum");</div></div>	<div><div><code></div><div>if (a &lt; b)</div><div>System.out.println(&quot;a is minimum&quot;);</div><div></code></div></div>

— М-да. Выглядит не очень красиво.

— А ты помнишь, в Java тоже экранируются некоторые символы — например «\». И тоже для его написания в строке приходится писать его дважды. Так что это часто встречаемое явление.

— Ок.

— У меня на сегодня все.

−

+82

+

Грицуник Андрей

Уровень 50, Москва, Россия

4 декабря 2021, 23:35

...

[Видео про XML и не только](#)

Ответить

-

+2

+

PaiMei in J#

Grand Master в Eagles' Claw

25 октября 2021, 15:19

...

Оказывается вот этот знак '&' называется амперсанд, для меня всегда он был and - загагулиной

Ответить

-

0

+

LuneFox

инженер по сопровождению в BIFIT

EXPERT

3 января, 18:48

...

Сделай локализованную версию XML и введи последовательность &заг; для загагулины :)

Ответить

-

+1

+

Вадим Илларионович

Java Developer в ЛАНИТ

24 марта 2021, 02:38

...

Подскажите пожалуйста с каким jdk работаете?
Новую версию установили уже?
Или работать на 15ой?

Ответить

-

0

+

Darth Nihilus

Разработчик спокойствия в Rage&Flame Industrie

21 мая 2021, 02:39

...

12)

Ответить

-

0

+

Anonymous #2489173

Уровень 35

20 сентября 2021, 11:17

...

15? 12? что???
разве тут до сих пор не 7 или 8?

Ответить

-

0

+

alex_us

Уровень 41, Симферополь

11 января 2021, 14:38

...

<https://javarush.ru/groups/posts/620-konkurs-osnovih-xml-dlja-java-programmista---chastjh-1-iz-3>
Почитайте эту ссылку и далее .Хорошо объясняются общие принципы. Что как и зачем

Ответить

-

+10

+

barracuda

Уровень 41, Санкт-Петербург, Россия

EXPERT

3 февраля 2021, 20:28

...

погрузился и охренел...

Ответить

-

0

+

Илья

Backend Developer в СберТех

6 февраля 2021, 11:55

...

как-то прочитал эту статью с пару недель назад, так после нее эта лекция зашла на раз два

Ответить

-

+1

+

Любовь

Уровень 33, Санкт-Петербург

27 апреля 2021, 13:44

...

спасибо, очень полезные статьи, все разложено по полочкам

Ответить

-

0

+

Антон

Full Stack Developer

1 декабря 2020, 22:03

...

А как же хранение символов юникода?

Ответить

-

0

+

Андрей

Уровень 41, Москва, Россия

29 ноября 2020, 18:13

...

JSON явно проще. Тут груда какого-то хлама с ссылками куда-то зачем-то. Самое интересно - так и не ясно, зачем нужен XML, когда есть JSON? (хотя XML появился раньше, но здесь выдано наоборот)

Ответить

-

+1

+

MKIV

Software Developer

21 декабря 2020, 20:42

...

"зачем нужен XML, когда есть JSON?" - я вижу преимущество XML в том, что он обязан подчиняться схеме (файл должен быть валидным). Структура XML чаще читается легче чем json.

Ответить

-

+1

+

Павел Павличенко

Уровень 35, Львов

22 августа 2020, 16:49

...

С каких пор Диего такой добрый? Так долго держал марку мудака, а теперь размяк

Ответить

-

+4

+

Евгений

Ведущий инженер в ПАО Сбербанк

EXPERT

16 июня 2020, 20:25

Важно понимать, что ссылка в xmlns - это не ссылка, а просто текст. Просто он должен быть уникальным, и потому принято писать туда URI.

Вместо

1

xmlns:soap="http://cxf.apache.org/bindings/soap"

можно написать:

1

xmlns:soap="Wow-wow, polegche"

но во втором случае будет сложнее обеспечить уникальность такого текста.

Ответить

+7

PaiMei in J#

Grand Master в Eagles' Claw

25 октября 2021, 15:30

Почему-то мне кажется, что как раз второй пример обеспечит лучшую уникальность, чем первый, затрудняюсь представить себе разработчика, который повторит конкретный xmlns ("Wow-wow, polegche") в конкретном проекте, таким образом, что это вызовет конфликт имен

Ответить

0

Евгений

Ведущий инженер в ПАО Сбербанк

EXPERT

26 октября 2021, 21:37

Ну сейчас как раз углублённо изучаю jaxb, так вот писать uri - это такое требование. Можно писать адрес своего сайта, он-то будет уникален скорее всего)

Ответить

0

PaiMei in J#

Grand Master в Eagles' Claw

27 октября 2021, 12:05

Подскажите, пожалуйста, какие источники используете при изучении XML? МБ литература / youtube-канал или что-нибудь в этом роде, буду крайне признателен

Ответить

0

Евгений

Ведущий инженер в ПАО Сбербанк

EXPERT

27 октября 2021, 18:41

А там особо нечего изучать, он на то и xml, что он очень гибкий. Я сейчас прохожу курс masterJava от Javaops, там немного xml разбирается. Ну и из полезного только вот такая статья есть <http://genberm.narod.ru/xml/lections/namespace/namespace.html>

Ответить

0

PaiMei in J#

Grand Master в Eagles' Claw

28 октября 2021, 15:06

Благодарю)

Ответить

0

Alex T

Уровень 35, Москва, Россия

25 апреля 2020, 20:55

знаете, меня часто добивают эти сокращения, которые ну казалось бы не несут никакой смысловой нагрузки(в большинстве случаев). Как пальцем в алфавит ткнули. Приходится запоминать. Но как только понимаешь смысл, то уже и запоминать не приходится - все становится логичным. Как тот же XML, хорошо написали, как расшифровывается.
те же < ?gt; - почему именно это сочетание букв. Снова надо запоминать?
Heee, вот - <https://stackoverflow.com/questions/5068951/what-do-lt-and-gt-stand-for>

Ответить

+17

miner

satellite engineer в AO VISAT-TEL

26 мая 2020, 08:13

thanks you
Less then
Greater then

Ответить

+2

PaiMei in J#

Grand Master в Eagles' Claw

25 октября 2021, 15:17

lt - less thAn
gt - greater thAn
amp - ampersand (да-да, вот эта штука '&' так и называется)
apos - apostrophe
quot - quotation mark

Ответить

+2

Показать еще комментарии



JavaRush — это интерактивный онлайн-курс по изучению Java-программирования с нуля. Он содержит 1200 практических задач с проверкой решения в один клик, необходимый минимум теории по основам Java и мотивирующие фишки, которые помогут пройти курс до конца: игры, опросы, интересные проекты и статьи об эффективном обучении и карьере Java-девелопера.

ПОДПИСЫВАЙТЕСЬ

ЯЗЫК ИНТЕРФЕЙСА

 Русский

▼

СКАЧИВАЙТЕ НАШИ ПРИЛОЖЕНИЯ

 ДОСТУПНО В
Google Play

 Загрузите в
App Store

