

Санкт-Петербургский национальный исследовательский университет  
информационных технологий, механики и оптики

Факультет инфокоммуникационных технологий

Направление подготовки 11.03.02

Лабораторная работа №4

«Работа с DTD схемами документов»

Выполнил:

Швалов Даниил Андреевич

Группа: К33211

Проверила:

Марченко Елена Вадимовна

Санкт-Петербург

2024

## 1. Введение

**Цель работы:** создание DTD-схем для проверки корректности различных XML-документов.

## 2. Ход работы

### Упражнение №1. Создание DTD схемы документа

В этом упражнении необходимо создать DTD-схему для документа *lessons.xml* из прошлой лабораторной работы. Содержимое документа *lessons.xml* представлено на рисунке 1. В данном документе содержатся следующие сущности:

- **lessons**: сущность, представляющая список дисциплин.
- **lesson**: сущность, представляющая конкретную дисциплину. Содержит название дисциплины, фамилию преподавателя, номер семестра, количество академических часов, тип контроля и формат проведения. Также обладает уникальным идентификатором.
- **CourseName**: сущность, представляющая название дисциплины.
- **ProfessorSurname**: сущность, представляющая имя преподавателя, который проводит данную дисциплину.
- **Semester**: сущность, представляющая номер семестра, в котором проводится дисциплина.
- **AcademicHours**: сущность, представляющая количество академических часов данной дисциплины.
- **ControlType**: сущность, представляющая тип контроля (например, экзамен или зачёт).
- **Format**: сущность, представляющая формат проведения дисциплины (например, очный или дистанционный).

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE lessons SYSTEM "lessons_DTD.dtd">
<lessons>
  <lesson id="m1">
    <CourseName>Математика</CourseName>
    <ProfessorSurname>Кольцова</ProfessorSurname>
    <Semester>1</Semester>
    <AcademicHours>108</AcademicHours>
    <ControlType>Экзамен</ControlType>
    <Format>Очный</Format>
  </lesson>
  <lesson id="m2">
    <CourseName>Программирование</CourseName>
    <ProfessorSurname>Осипов</ProfessorSurname>
    <Semester>2</Semester>
    <AcademicHours>144</AcademicHours>
    <ControlType>Зачёт</ControlType>
    <Format>Очно-дистанционный</Format>
  </lesson>
  <lesson id="m3">
    <CourseName>Базы данных</CourseName>
    <ProfessorSurname>Осетрова</ProfessorSurname>
    <Semester>3</Semester>
    <AcademicHours>96</AcademicHours>
    <ControlType>Дифференцированный зачёт</ControlType>
    <Format>Дистанционный</Format>
  </lesson>
</lessons>

```

Рисунок 1 – Содержимое документа *lessons.xml*

Исходя из ранее приведенного описания сущностей была сделана DTD-схема, приведенная на рисунке 2. Для идентификаторов дисциплин используется тип *ID*, который проверяет уникальность идентификатора. Для конченных значений, таких как название дисциплины, используется тип *PCDATA*, который позволяет хранить в себе данные, которые будут интерпретированы как разметка XML.

```

<!ELEMENT lessons (lesson+)>
<!ELEMENT lesson (
    CourseName,
    ProfessorSurname,
    Semester,
    AcademicHours,
    ControlType,
    Format
)>
<!ATTLIST lesson id ID #REQUIRED>
<!ELEMENT CourseName (#PCDATA)>
<!ELEMENT ProfessorSurname (#PCDATA)>
<!ELEMENT Semester (#PCDATA)>
<!ELEMENT AcademicHours (#PCDATA)>
<!ELEMENT ControlType (#PCDATA)>
<!ELEMENT Format (#PCDATA)>

```

Рисунок 2 – Содержимое схемы *lessons\_DTD.dtd*

## Упражнение №2. Работа с DTD-схемой

В данном упражнении необходимо для существующей DTD-схемы создать XML-документ, который будет удовлетворять этой схеме. В качестве DTD-схемы был дан файл, показанный на рисунке 3. В нем содержатся следующие сущности:

- **library**: представляет собой библиотеку, хранит информацию о книгах и авторах.

- **book\_catalog**: представляет собой каталог книг, хранит информацию о всех книгах.

- **book**: представляет собой книгу, хранит информацию об авторах книги, ее заголовок, издательство и аннотацию к книге. Имеет такие атрибуты, как идентификатор, ISBN, год издания и тип (перевод или оригинал).

- **authors**: представляет собой авторов книги.

- **author**: представляет собой конкретного автора книги.

- **title**: представляет собой заголовок книги.

- **publishing**: представляет собой издательство.

- **annotation**: представляет собой аннотацию к книге.
- **author\_catalog**: представляет собой каталог авторов, хранит информацию о книгах авторов.
- **author\_book**: представляет собой книгу автора. Имеет ссылку на идентификатор книги.

```

<!ELEMENT library (book_catalog, author_catalog)>

<!ELEMENT book_catalog (book*)>
<!ELEMENT book (authors?, title, publishing, annotation?)>
<!ELEMENT authors (author+)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT publishing (#PCDATA)>
<!ELEMENT annotation (#PCDATA)>

<!ELEMENT author_catalog (author_book*)>
<!ELEMENT author_book (#PCDATA)>

<!ATTLIST book
  id ID #REQUIRED
  isbn CDATA #IMPLIED
  year CDATA #REQUIRED
  issue (translation | original) "translation">

<!ATTLIST author_book idref IDREF #REQUIRED>

<!ENTITY linux "Linux">
<!ENTITY internet "Internet">

```

Рисунок 3 – Содержимое схемы *library\_DTD.dtd*

Исходя из ранее описанных атрибутов, был составлен XML-документ, приведенный на рисунке 4. В нем приведены три книги, каждая из которых по-разному использует различные сущности. Так, первая сущность содержит всевозможные атрибуты и поля. Вторая сущность показывает, что книга может иметь несколько авторов. Третья сущность практически не обладает полями и демонстрирует работу

необязательных полей. Также в документе созданы две ссылки авторов на конкретные книги.

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<!DOCTYPE library SYSTEM "library_DTD.dtd">
```

```
<library>
```

```
  <book_catalog>
```

```
    <book id="m1" isbn="978-5-17-090630-7" year="2024" issue="original">
```

```
      <authors>
```

```
        <author>Достоевский Ф. М.</author>
```

```
      </authors>
```

```
      <title>Преступление и наказание</title>
```

```
      <publishing>ACE</publishing>
```

```
      <annotation>
```

"Преступление и наказание" (1866) — одно из самых значительных произведений в истории мировой литературы. Это и глубокий философский роман, и тонкая психологическая драма, и захватывающий детектив, и величественная картина мрачного города, в недрах которого герои грешат и ищут прощения, жертвуют собой и отрекаются от себя ради близких и находят успокоение в смирении, покаянии, вере. Главный герой романа Родион Раскольников решается на убийство, чтобы доказать себе и миру, что он не "тварь дрожащая", а "право имеет". Главным предметом исследования писателя становится процесс превращения добропорядочного, умного и доброго юноши в убийцу, а также то, как совершивший преступление Раскольников может искупить свою вину.

```
    </annotation>
```

```
  </book>
```

```
  <book id="m2" isbn="978-5-4461-2024-6" year="2023">
```

```
    <authors>
```

```
      <author>Форд Нил</author>
```

```
      <author>Ричардс Марк</author>
```

```
    </authors>
```

```
    <title>
```

Современный подход к программной архитектуре: сложные компромиссы

```
  </title>
```

```
  <publishing>Питер</publishing>
```

```
  <annotation>
```

В архитектуре программного обеспечения нет простых решений. Напротив, есть масса сложностей — задач и проблем, для решения которых нет готовых ответов и приходится выбирать между различными компромиссами. Эта книга научит вас критически относиться к компромиссам, связанным с

```

        распределенными архитектурами.
    </annotation>
</book>
<book id="m3" year="999" issue="original">
    <title>Редкая книга</title>
    <publishing>Неизвестное издательство</publishing>
</book>
</book_catalog>
<author_catalog>
    <author_book idref="m1"/>
    <author_book idref="m2"/>
</author_catalog>
</library>

```

Рисунок 4 – Содержимое документа *library.xml*

### Упражнение №3. Создание DTD-схемы

В предыдущей лабораторной работе был создан XML-документ *books.xml*, содержимое которого представлено на рисунке 5. В нем содержатся следующие сущности:

- **books**: сущность, представляющая список всех книг.
- **book**: сущность, представляющая конкретную книгу. Содержит заголовок, аннотацию, автора, жанр, количество страниц, язык издания, издательство, год издания, ISBN и возрастное ограничение. Также имеет уникальный идентификатор.
- **title**: сущность, представляющая заголовок книги.
- **summary**: сущность, представляющая аннотацию к книге.
- **authors**: сущность, представляющая авторов книги.
- **author**: сущность, представляющая конкретного автора книги.
- **genres**: сущность, представляющая жанры книги.
- **genre**: сущность, представляющая конкретный жанр книги.
- **pages**: сущность, представляющая количество страниц.
- **language**: сущность, представляющая язык, на котором издана книга.

- **publisher**: сущность, представляющая издательство.
- **publication\_year**: сущность, представляющая год издания книги.
- **isbn**: сущность, представляющая ISBN книги.
- **age\_limit**: сущность, представляющая возрастное ограничение книги.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE books SYSTEM "books_DTD.dtd">
```

```
<books>
```

```
  <book id="m1">
```

```
    <title>Ненасильственное общение. Язык Жизни</title>
```

```
    <summary>
```

Метод ненасильственного общения (ННО) реально улучшает жизнь тысяч людей. Он применим и в супружеской спальне, и в классной комнате, и за столом переговоров, и на линии фронта. Никогда не поздно отказаться от насилия во благо всех заинтересованных сторон. Основатель метода ННО Маршалл Розенберг подробно объясняет его суть и на многочисленных примерах из своей практики миротворца показывает, как им пользоваться. В новом издании книги добавлена глава Разрешение конфликтов и посредничество.

```
  </summary>
```

```
  <authors>
```

```
    <author>Розенберг Маршалл</author>
```

```
  </authors>
```

```
  <genres>
```

```
    <genre>Психология</genre>
```

```
  </genres>
```

```
  <pages>288</pages>
```

```
  <language>RU</language>
```

```
  <publisher>СОФИЯ МЕДИА</publisher>
```

```
  <publication_year>2020</publication_year>
```

```
  <isbn>978-5-90-689777-0</isbn>
```

```
  <age_limit>16</age_limit>
```

```
</book>
```

```
<book id="m2">
```

```
  <title>Ведьмак. Последнее желание</title>
```

```
  <summary>
```

Ведьмак — это мастер меча и мэтр волшебства, ведущий непрерывную войну с кровожадными монстрами, которые угрожают покою сказочной страны. «Ведьмак» — это мир на острие меча, ошеломляющее действие, незабываемые ситуации, великолепные боевые сцены.



```

</summary>
<authors>
  <author>Анджей Сапковский</author>
</authors>
<genres>
  <genre>Фэнтези</genre>
</genres>
<pages>320</pages>
<language>RU</language>
<publisher>Neoclassic</publisher>
<publication_year>2022</publication_year>
<isbn>978-5-17-073052-0</isbn>
<age_limit>16</age_limit>
</book>
<book id="m3">
  <title>Harry Potter and the Philosophers Stone</title>
  <summary>
    Harry Potter has never even heard of Hogwarts when the letters start
    dropping on the doormat at number four, Privet Drive. Addressed in green
    ink on yellowish parchment with a purple seal, they are swiftly
    confiscated by his grisly aunt and uncle. Then, on Harry's eleventh
    birthday, a great beetle-eyed giant of a man called Rubeus Hagrid bursts
    in with some astonishing news: Harry Potter is a wizard, and he has a
    place at Hogwarts School of Witchcraft and Wizardry.
  </summary>
  <authors>
    <author>J. K. Rowling</author>
  </authors>
  <genres>
    <genre>Фэнтези</genre>
  </genres>
  <pages>342</pages>
  <language>EN</language>
  <publisher>Bloomsbury</publisher>
  <publication_year>2014</publication_year>
  <isbn>978-1-40-885565-2</isbn>
  <age_limit>12</age_limit>
</book>
</books>

```

## Рисунок 5 – Содержимое документа *books.xml*

Исходя из выше описанных сущностей, была создана DTD-схема, представленная на рисунке 6.

```
<!ELEMENT books (book*)>

<!ELEMENT book (
    title,
    summary?,
    authors?,
    genres,
    pages,
    language,
    publisher?,
    publication_year?,
    isbn?,
    age_limit
)>
<!ATTLIST book id ID #REQUIRED>

<!ELEMENT authors (author*)>
<!ELEMENT genres (genre*)>

<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT summary (#PCDATA)>
<!ELEMENT genre (#PCDATA)>
<!ELEMENT pages (#PCDATA)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT publication_year (#PCDATA)>
<!ELEMENT isbn (#PCDATA)>
<!ELEMENT age_limit (#PCDATA)>
```

## Рисунок 6 – Содержимое схемы *books\_DTD.dtd*

### **3. Вывод**

В ходе выполнения лабораторной работы были созданы DTD-схемы для проверки корректности XML-документов.