

Тема №4 Java & XML

На данном занятии необходимо познакомиться с правилами создания *XML* – документов и приемами проверки их на корректность, изучить возможности *SAX* и *DOM* парсеров, рассмотреть, как с их помощью можно прочесть, изменить и сохранить *XML* – документ.

Основные задания

Задание №1

Дан зашумленный набор некоторых экспериментально полученных данных (x, y) , которые располагаются на графике примерно вдоль прямой линии (см. *Рис. 4.1*).

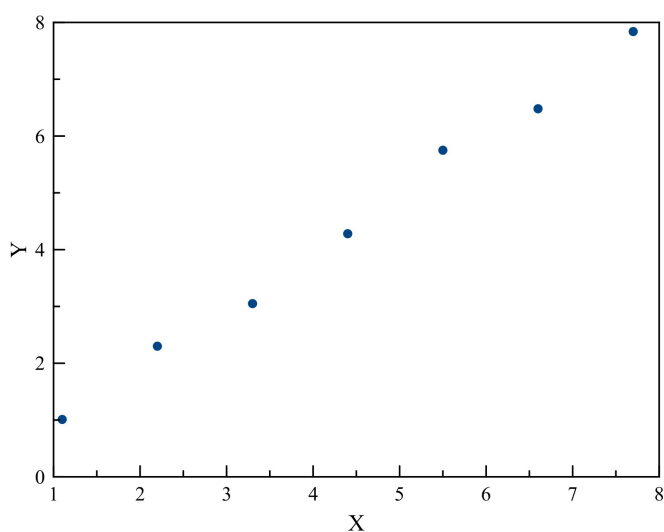


Рисунок 4.1. Экспериментально полученные данные

Вычислить коэффициенты прямой $y = k \cdot x + b$, которая наилучшим образом приближает результаты.

Для проверки правильности работы программы использовать приведенные ниже данные, хранящиеся в файле *XML*–формата:

x	y	Дата наблюдения
1.1	1.01	19.03.2018
2.2	2.30	20.03.2018
3.3	3.05	21.03.2018
4.4	4.28	22.03.2018
5.5	5.75	23.03.2018
6.6	6.48	24.03.2018
7.7	7.84	25.03.2018

Теоретические сведения

Самый простой и наиболее часто используемый вид регрессии (зависимости среднего значения какой-либо случайной величины от другой величины) — линейная. Приближение данных (x_i, y_i) осуществляется линейной функцией $y(x) = k \cdot x + b$. Если определить коэффициенты регрессии k и b , то для любого значения x можно предсказать значение y , просто проведя вычисления по формуле. Стандартный метод получения коэффициентов k и b — это метод наименьших квадратов. Метод получил такое название, поскольку коэффициенты k и b вычисляются из условия минимизации суммы квадратов ошибок $|b + k \cdot x_i - y_i|^2$. Результирующие расчетные формулы можно записать так:

$$k = \frac{(\sum x \cdot y) - (\sum x) \cdot \bar{y}}{(\sum x^2) - (\sum x) \cdot \bar{x}}, \quad b = \bar{y} - k \bar{x}$$

здесь суммирование проводится по всем данным, а \bar{x} , \bar{y} — средние значения соответствующих величин.

Задания для работы

1. Создать *XML*-файл, соответствующий указанному набору данных и проверить его корректность.
2. Создать приложение для разбора полученного *XML*-файла с данными с помощью *SAX*-парсера без проверки действительности документа (*without validation*). Необходимые вычисления по указанным выше формулам провести «на лету», непосредственно во время чтения документа.
3. Для *XML*-файла с данными создать файлы с *DTD* и *XSD* схемами для проверки действительности документа (*validation*) и проверить правильность их написания.
4. Создать приложение для разбора *XML*-файла с данными при помощи *SAX*-парсера с проверкой действительности документа (*with validation*) с применением *DTD*, *XSD* схем. Во время чтения и разбора *XML*-файла построить в памяти адекватную структуру данных для хранения прочитанных значений; создать класс для анализа прочитанных данных.
5. Создать приложение для разбора *XML*-файла с данными при помощи *DOM*-парсера без проверки и с проверкой действительности документа (*without and with validation*) с применением *DTD* и *XSD* схем. Для чтения с проверкой действительности документа создать класс-оболочку, предназначенную для работы с *DOM*-объектом, как со структурой данных с сохраненными значениями. Предусмотреть возможность изменять значения, хранящиеся в *DOM*-узлах, добавлять, удалять, вставлять и заменять *DOM*-узлы. Сохранить измененный *DOM*-объект в новый *XML*-файл.
6. Создать приложение для разбора *XML*-файла с данными с помощью *SAX*-парсера с проверкой действительности документа (*with validation*) с применением *DTD* и *XSD* схем; построить в памяти адекватную структуру данных для хранения прочитанных значений; создать класс для анализа прочитанных данных. Преобразовать полученную структуру данных и вычисленный результат в *DOM*-объект и сохранить его в новый *XML*-файл.

Новый, результирующий *XML*-файл может иметь такую структуру:

```
<?xml version="1.0" encoding="Windows-1251" standalone="no"?>
<Analyser>
<dataTable>
<dataPoint date="16.03.2018">
<x>1.1</x>
```