

Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Лабораторная работа №4

**Исследование протоколов, форматов обмена информацией и
языков разметки документов**

Вариант №23

Выполнил

Макогон Ярослав Вадимович

Номер группы: Р3118

Проверила

Малышева Т. А.

Содержание

Задание.....	3
Основные этапы вычислений.....	6
Заключение.....	14
Список использованных источников	15

Задание

Обязательное задание:

1. Определить номер варианта как остаток деления на 36 последних двух цифр своего идентификационного номера в ISU: например, $125598 / 36 = 26$. В случае, если в оба указанных дня недели нет занятий, то увеличить номер варианта на восемь. В случае, если занятий нет и в новом наборе дней, то продолжать увеличивать на восемь.
2. Изучить форму Бэкуса-Наура.
3. Изучить основные принципы организации формальных грамматик.
4. Изучить особенности языков разметки/форматов JSON, YAML, XML.
5. Понять устройство страницы с расписанием на примере расписания лектора: https://itmo.ru/ru/schedule/3/125598/raspisanie_zanyatiy.htm
6. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы хотя бы в одной из выбранных дней было не менее двух занятий (можно использовать своё персональное). В случае, если в данный день недели нет таких занятий, то увеличить номер варианта ещё на восемь.
7. Обязательное задание (позволяет набрать до 45 процентов от максимального числа баллов БаРС за данную лабораторную): написать программу на языке Python 3.x или любом другом, которая бы осуществляла парсинг и конвертацию исходного файла в новый путём простой замены метасимволов исходного формата на метасимволы результирующего формата.
8. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

Дополнительное задание №1:

- a) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
- b) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать. 1
- c) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №2:

- a) Переписать исходный код, добавив в него использование регулярных выражений.
- b) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №3:

- a) Переписать исходный код таким образом, чтобы для решения задачи использовались формальные грамматики. То есть ваш код должен уметь осуществлять парсинг и

конвертацию любых данных, представленных в исходном формате, в данные, представленные в результирующем формате: как с готовыми библиотеками из дополнительного задания №1.

- b) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
- c) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №4:

- a) Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.
- b) Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №5:

- a) Переписать исходную программу, чтобы она осуществляла парсинг и конвертацию исходного файла в любой другой формат (кроме JSON, YAML, XML, HTML): PROTOBUF, TSV, CSV, WML и т.п. 2 (**XAML**)
- b) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

- Проверить, что все пункты задания выполнены и выполнены верно.
- Написать отчёт о проделанной работе.
- Подготовиться к устным вопросам на защите.

23	XML	YAML	Вторник, четверг
----	-----	------	------------------

вт, 12 ноября

10:00 Лекции

11:30 Линейная алгебра
Рванова Алла
Ауд. ауд. 2426 (2426)
Кронверкский пр., д.49,
лит.А

11:40 Практические занятия

13:10 Линейная алгебра
Рванова Алла
Ауд.2430
Кронверкский пр., д.49,
лит.А

15:20 Практические занятия

16:50 Английский язык В1.2
Морозова Анна
Михайловна
Ауд.3307
ул.Ломоносова, д.9, лит. Е

17:00 Практические занятия

18:30 Английский язык В1.2
Морозова Анна
Михайловна
Ауд.3307
ул.Ломоносова, д.9, лит. Е

чт, 14 ноября

10:00 Практические занятия

11:30 История русской
культуры в контексте
мировой культуры
Щеглова Иванна
Павловна
Ауд.2412
Кронверкский пр., д.49,
лит.А

11:40 Лекции

13:10 История русской
культуры в контексте
мировой культуры
Жиркова Галина Петровна
Ауд.1404
Кронверкский пр., д.49,
лит.А

Основные этапы вычислений

Составлено расписание в формате XML:

```
<schedule>
  <day>
    <date>5 ноября</date>
    <dayOfWeek>вторник</dayOfWeek>
    <class>
      <time>10:00 - 11:30</time>
      <type>Лекция</type>
      <subject>Линейная алгебра</subject>
      <teacher>Рванова Алла</teacher>
      <location>ауд. 2426, Кронверкский пр., д.49, лит.А</location>
    </class>
    <class>
      <time>11:40 - 13:10</time>
      <type>Практическое занятие</type>
      <subject>Линейная алгебра</subject>
      <teacher>Рванова Алла</teacher>
      <location>ауд. 2430, Кронверкский пр., д.49, лит.А</location>
    </class>
    <class>
      <time>15:20 - 16:50</time>
      <type>Практическое занятие</type>
      <subject>Английский язык В1.2</subject>
      <teacher>Морозова Анна Михайловна</teacher>
      <location>ауд. 3307, ул. Ломоносова, д.9, лит. Е</location>
    </class>
    <class>
      <time>17:00 - 18:30</time>
      <type>Практическое занятие</type>
      <subject>Английский язык В1.2</subject>
      <teacher>Морозова Анна Михайловна</teacher>
      <location>ауд. 3307, ул. Ломоносова, д.9, лит. Е</location>
    </class>
  </day>
  <day>
    <date>14 ноября</date>
    <dayOfWeek>четверг</dayOfWeek>
    <class>
      <time>10:00 - 11:30</time>
      <type>Практическое занятие</type>
      <subject>История русской культуры в контексте мировой
культуры</subject>
      <teacher>Щеглова Иванна Павловна</teacher>
      <location>ауд. 2412, Кронверкский пр., д.49, лит.А</location>
    </class>
    <class>
      <time>11:40 - 13:10</time>
      <type>Лекция</type>
      <subject>История русской культуры в контексте мировой
культуры</subject>
      <teacher>Жиркова Галина Петровна</teacher>
      <location>ауд. 1404, Кронверкский пр., д.49, лит.А</location>
    </class>
  </day>
</schedule>
```

Обязательное задание:

Написана программа по заданным требованиям:

<https://github.com/Not-N0w/Informatics-lab4/blob/main/nec.py>

Результат работы (ввод: xml, вывод: yaml):

```
schedule:
  day:
    - date: 5 ноября
      dayOfWeek: вторник
      class:
        - time: 10:00 - 11:30
          type: Лекция
          subject: Линейная алгебра
          teacher: Рванова Алла
          location: ауд. 2426, Кронверкский пр., д.49, лит.А
        - time: 11:40 - 13:10
          type: Практическое занятие
          subject: Линейная алгебра
          teacher: Рванова Алла
          location: ауд. 2430, Кронверкский пр., д.49, лит.А
        - time: 15:20 - 16:50
          type: Практическое занятие
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
        - time: 17:00 - 18:30
          type: Практическое занятие
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
    - date: 14 ноября
      dayOfWeek: четверг
      class:
        - time: 10:00 - 11:30
          type: Практическое занятие
          subject: История русской культуры в контексте мировой культуры
          teacher: Щеглова Иванна Павловна
          location: ауд. 2412, Кронверкский пр., д.49, лит.А
        - time: 11:40 - 13:10
          type: Лекция
          subject: История русской культуры в контексте мировой культуры
          teacher: Жиркова Галина Петровна
          location: ауд. 1404, Кронверкский пр., д.49, лит.А
```

Дополнительное задание №1:

Написана программа по заданным требованиям:

<https://github.com/Not-Now/Informatics-lab4/blob/main/ext1.py>

Результат работы (ввод: xml, вывод: yaml):

```
schedule:
  day:
    - class:
        - location: ауд. 2426, Кронверкский пр., д.49, лит.А
          subject: Линейная алгебра
          teacher: Рванова Алла
          time: 10:00 - 11:30
          type: Лекция
        - location: ауд. 2430, Кронверкский пр., д.49, лит.А
          subject: Линейная алгебра
          teacher: Рванова Алла
          time: 11:40 - 13:10
          type: Практическое занятие
        - location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          time: 15:20 - 16:50
          type: Практическое занятие
        - location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          time: 17:00 - 18:30
          type: Практическое занятие
      date: 5 ноября
      dayOfWeek: вторник
    - class:
        - location: ауд. 2412, Кронверкский пр., д.49, лит.А
          subject: История русской культуры в контексте мировой культуры
          teacher: Щеглова Иванна Павловна
          time: 10:00 - 11:30
          type: Практическое занятие
        - location: ауд. 1404, Кронверкский пр., д.49, лит.А
          subject: История русской культуры в контексте мировой культуры
          teacher: Жиркова Галина Петровна
          time: 11:40 - 13:10
          type: Лекция
      date: 14 ноября
      dayOfWeek: четверг
```

Результат работы программы из обязательного задания практически идентичен результату работы программы из дополнительного задания 1. Различие лишь в порядке следования тегов, что может быть обусловлено лишь особенностями алгоритма парсинга.

Дополнительное задание №2:

Написана программа по заданным требованиям:

<https://github.com/Not-N0w/Informatics-lab4/blob/main/ext2.py>

Результат работы (ввод: xml, вывод: yaml):

```
schedule:
  day:
    - date: 5 ноября
      dayOfWeek: вторник
      class:
        - time: 10:00 - 11:30
          type: Лекция
          subject: Линейная алгебра
          teacher: Рванова Алла
          location: ауд. 2426, Кронверкский пр., д.49, лит.А
        - time: 11:40 - 13:10
          type: Практическое занятие
          subject: Линейная алгебра
          teacher: Рванова Алла
          location: ауд. 2430, Кронверкский пр., д.49, лит.А
        - time: 15:20 - 16:50
          type: Практическое занятие
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
        - time: 17:00 - 18:30
          type: Практическое занятие
          subject: Английский язык В1.2
          teacher: Морозова Анна Михайловна
          location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
      - date: 14 ноября
        dayOfWeek: четверг
        class:
          - time: 10:00 - 11:30
            type: Практическое занятие
            subject: История русской культуры в контексте мировой культуры
            teacher: Щеглова Иванна Павловна
            location: ауд. 2412, Кронверкский пр., д.49, лит.А
          - time: 11:40 - 13:10
            type: Лекция
            subject: История русской культуры в контексте мировой культуры
            teacher: Жиркова Галина Петровна
            location: ауд. 1404, Кронверкский пр., д.49, лит.А
```

Результат работы программы из обязательного задания идентичен результату работы программы из дополнительного задания 2. Различие между результатами работы программ из дополнительных заданий 1 и 2 заключается только в порядке следования тегов, что может быть обусловлено лишь особенностями алгоритма парсинга.

Дополнительное задание №3:

Написана программа по заданным требованиям:

<https://github.com/Not-N0w/Informatics-lab4/blob/main/ext3.py>

Результат работы (ввод: xml, вывод: yaml):

```
schedule:
- day:
  date: 5 ноября
  dayOfWeek: вторник
  class:
  - time: 10:00 - 11:30
    type: Лекция
    subject: Линейная алгебра
    teacher: Рванова Алла
    location: ауд. 2426, Кронверкский пр., д.49, лит.А
  - time: 11:40 - 13:10
    type: Практическое занятие
    subject: Линейная алгебра
    teacher: Рванова Алла
    location: ауд. 2430, Кронверкский пр., д.49, лит.А
  - time: 15:20 - 16:50
    type: Практическое занятие
    subject: Английский язык В1.2
    teacher: Морозова Анна Михайловна
    location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
  - time: 17:00 - 18:30
    type: Практическое занятие
    subject: Английский язык В1.2
    teacher: Морозова Анна Михайловна
    location: ауд. 3307, ул. Ломоносова, д.9, лит. Е
- date: 14 ноября
  dayOfWeek: четверг
  - time: 10:00 - 11:30
    type: Практическое занятие
    subject: История русской культуры в контексте мировой культуры
    teacher: Щеглова Иванна Павловна
    location: ауд. 2412, Кронверкский пр., д.49, лит.А
  - time: 11:40 - 13:10
    type: Лекция
    subject: История русской культуры в контексте мировой культуры
    teacher: Жиркова Галина Петровна
    location: ауд. 1404, Кронверкский пр., д.49, лит.А
```

Результат работы программы из обязательного задания и дополнительного задания 2 идентичен результату работы программы из дополнительного задания 3, за исключением “-” на второй строке. Это возникает из-за особенности написанного рекурсивного спуска: мы не можем предвидеть количество дочерних тегов с одинаковым названием. Несмотря на это, оба вывода являются верными для заданного xml. Отличия с 1 доп. заданием и обязательным заключаются еще и в порядке следования вложенных строк, что может быть обусловлено лишь особенностями алгоритма парсинга.

Дополнительное задание №4:

Написана программа по заданным требованиям:

<https://github.com/Not-N0w/Informatics-lab4/blob/main/ext4.py>

Результат работы (ввод: xml, вывод: yaml):

```
Время выполнения парсинга и конвертации (x100) в обязательном задании: 0.0465846061706543
Время выполнения парсинга и конвертации (x100) в дополнительном задании 1: 0.20058369636535645
Время выполнения парсинга и конвертации (x100) в дополнительном задании 2: 0.16306543350219727
Время выполнения парсинга и конвертации (x100) в дополнительном задании 3: 0.053941965103149414
```

Рисунок 1. Результат работы программы из дополнительного задания 4

Самым быстрым оказался способ конвертации в обязательном задании, потому что в нем используется неплохая механика замены, которая хорошо работает на небольших примерах вроде расписания. На втором месте по скорости – рекурсивный спуск из доп. задания 4, который так же отлично работает на небольших файлах. Далее идет конвертация при помощи регулярных выражений. Они не супер быстро работают (хотя и неплохо оптимизированы в Python) + особенности конкретно этой идеи конвертации так же повлияли на скорость работы программы. Программа, написанная при помощи сторонних библиотек, оказалась самой медленной, потому что при ее работе происходит сначала конвертация в промежуточный формат, а затем уже в необходимый.

Дополнительное задание №5:

Написана программа по заданным требованиям:

<https://github.com/Not-Now/Informatics-lab4/blob/main/ext5.py>

Результат работы (ввод: xml, вывод: xaml):

```
<schedule>
  <day>
    <date text = '5 ноября' />
    <dayOfWeek text = 'вторник' />
    <class>
      <time text = '10:00 - 11:30' />
      <type text = 'Лекция' />
      <subject text = 'Линейная алгебра' />
      <teacher text = 'Рванова Алла' />
      <location text = 'ауд. 2426, Кронверкский пр., д.49, лит.А' />
    </class>
    <class>
      <time text = '11:40 - 13:10' />
      <type text = 'Практическое занятие' />
      <subject text = 'Линейная алгебра' />
      <teacher text = 'Рванова Алла' />
      <location text = 'ауд. 2430, Кронверкский пр., д.49, лит.А' />
    </class>
    <class>
      <time text = '15:20 - 16:50' />
      <type text = 'Практическое занятие' />
      <subject text = 'Английский язык В1.2' />
      <teacher text = 'Морозова Анна Михайловна' />
      <location text = 'ауд. 3307, ул. Ломоносова, д.9, лит. Е' />
    </class>
    <class>
      <time text = '17:00 - 18:30' />
      <type text = 'Практическое занятие' />
      <subject text = 'Английский язык В1.2' />
      <teacher text = 'Морозова Анна Михайловна' />
      <location text = 'ауд. 3307, ул. Ломоносова, д.9, лит. Е' />
    </class>
  </day>
  <day>
    <date text = '14 ноября' />
    <dayOfWeek text = 'четверг' />
    <class>
      <time text = '10:00 - 11:30' />
      <type text = 'Практическое занятие' />
      <subject text = 'История русской культуры в контексте мировой
культуры' />
      <teacher text = 'Щеглова Иванна Павловна' />
      <location text = 'ауд. 2412, Кронверкский пр., д.49, лит.А' />
    </class>
    <class>
      <time text = '11:40 - 13:10' />
      <type text = 'Лекция' />
      <subject text = 'История русской культуры в контексте мировой
культуры' />
      <teacher text = 'Жиркова Галина Петровна' />
```

```
        <location text = 'ауд. 1404, Кронверкский пр., д.49, лит.А' />
    </class>
</day>
</schedule>
```

XAML – язык разметки (UI) от Microsoft.

Особенности XAML:

- В XAML объекты, содержащие только текст помещаются в атрибут одиночного тега
- Поддерживает сложные элементы управления и контейнеры, например, для создания привязок и обработки событий.
- Многие свойства задаются как атрибуты элементов, но они могут быть сложными значениями или ссылками на другие элементы.
- Содержит элементы, специфичные для UI-приложений, такие как <Button>, <TextBlock>, <Window>, <Grid> и другие.

Заключение

- Изучил форму Бэкуса-Наура.
- Изучил основные принципы организации формальных грамматик.
- Изучил особенности языков разметки/форматов JSON, YAML, XML
- Написаны программы, выполняющие поставленные в лабораторной работе задания, что помогло закрепить изученный материал.

Список использованных источников

1. Статья “YAML” – URL: <https://blog.skillfactory.ru/glossary/yaml/>
2. Статья “YAML за 5 минут: синтаксис и основные возможности” – URL: <https://tproger.ru/translations/yaml-za-5-minut-sintaksis-i-osnovnye-vozmozhnosti>
3. Статья “ Обзор XAML (WPF .NET)” – URL: <https://learn.microsoft.com/ru-ru/dotnet/desktop/wpf/xaml/?view=netdesktop-8.0>