

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНОЙ ТЕХНИКИ

НАПРАВЛЕНИЕ ПРОГРАММНАЯ ИНЖЕНЕРИЯ

ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА СИСТЕМНОЕ И ПРИКЛАДНОЕ
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ № 4

курса «Информатика»

Вариант № 129

Выполнил студент:

Бых Даниил Максимович

группа: Р3109

Проверил:

Малышева Т. А.

Санкт-Петербург, 2025

Содержание

1. Задание	2
2. Основные этапы вычислений	4
1. Обязательное задание	4
2. Дополнительные задание 1	10
3. Дополнительные задание 2	10
4. Дополнительные задание 3	12
5. Дополнительные задание 4	13
3. Вывод	15
4. Литература	16

1. Задание

1. Определить номер варианта как остаток деления на 132 своего идентификационного номера в ISU: например, $125598 / 132 = 16$. В случае, если в обоих указанных днях 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, RON, HCL, YAML, TOML,INI, 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 50 процентов от максимального числа баллов БаРС за данную лабораторную). Написать программу на языке Python 3.x или любом другом, которая:
 - осуществляет парсинг и конвертацию исходного файла в бинарный объект (=десериализацию);
 - для решения задачи использует 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 №2);
 - не использует готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки файлов.
8. Дополнительное задание №1 (позволяет набрать +15 процентов от максимального числа баллов БаРС за данную лабораторную). Написать программу на языке Python 3.x или любом другом, которая:
 - осуществляет парсинг и конвертацию бинарного объекта, полученного в обязательном задании, в новый формат (=серIALIZАЦИЮ);
 - для решения задачи использует формальные грамматики;

- не использует готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки файлов.
9. Дополнительное задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
- Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов (десериализацию и сериализацию).
 - Переписать исходный код и код из дополнительного задания №1, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
 - Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
10. Дополнительное задание №3 (позволяет набрать +20 процентов от максимального числа баллов БаРС за данную лабораторную). Переписать код из дополнительного задания №1, чтобы сериализация происходила в XML файл.
11. Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).
- Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.
 - Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте
12. Проверить, что все пункты задания выполнены и выполнены верно
13. Написать отчёт о проделанной работе.
14. Подготовиться к устным вопросам на защите.

2. Основные этапы вычислений

2. 1. Обязательное задание

Так как мой номер ISU - 501993, то итоговый вариант $501993 \% 132 = 129$ (129 вариант). Необходимо было выполнить следующие преобразования:

- Перевести данные из файла в формате INI в файл в формате RON.
- В качестве дней недели взять понедельник и субботу

Код, являющийся точкой входа в программу для решения как обязательного, так и дополнительных заданий, находится в файле 1:

```
1 from lib.advanced_convertor import AdvancedConvertor
2 from lib.hand_written_convertor import HandWrittenConvertor
3 from lib.formal import BinarySerializer, Wrapper
4 from lib.test import TestRunner
5 from os import path
6
7
8 if __name__ == "__main__":
9     INPUT_FILE_SRC = path.abspath(path.join(path.dirname('__file__'), 'input', 'schedule.ini'))
10    BIN_FILE_PATH = path.abspath(path.join(path.dirname('__file__'), 'output', 'schedule_hand_written.bin'))
11
12    print(f"Вариант: {501993 % 132}\n-----|")
13
14    # Обязательное задание: десериализация в бинарный объект
15    ini_file_content = HandWrittenConvertor.read_file(INPUT_FILE_SRC)
16    serialized_raw = HandWrittenConvertor.deserialize(file_content=ini_file_content)
17    BinarySerializer.save_to_file(serialized_raw, BIN_FILE_PATH)
18
19    # Дополнительное задание 1: загрузка из бинарного + сериализация в RON
20    loaded_obj = BinarySerializer.load_from_file(BIN_FILE_PATH)
21    HandWrittenConvertor.serialize(obj=loaded_obj, format='ron')
22
23    # Дополнительное задание 2
24    ini_parser = AdvancedConvertor.read_file(INPUT_FILE_SRC)
25    serialized_advanced = AdvancedConvertor.deserialize(parse_object=ini_parser)
26    AdvancedConvertor.serialize(serialized_advanced, format='ron')
27
28    # Дополнительное задание 3: загрузка из бинарного + сериализация в XML
29    loaded_obj_xml = BinarySerializer.load_from_file(BIN_FILE_PATH)
30    HandWrittenConvertor.serialize(obj=loaded_obj_xml, format='xml')
31
32    # Дополнительное задание 4
33    TestRunner.run_handwritten_test(INPUT_FILE_SRC, BIN_FILE_PATH, 100)
34    TestRunner.run_advanced_test(INPUT_FILE_SRC)
```

Листинг 1: main.py

Разберем детально решение обязательного задания.

- для парсинга и конвертацию исходного файла в бинарный объект был создан класс HandWrittenConvertor в файле [2](#).
- в нем был создан метод read_file, осуществляющий чтение исходного файла и удаление лишних символов табуляции.
- также в нем был создан метод deserialize, осуществляющий десериализацию исходного файла в объект класса Wrapper, а затем в бинарный файл при помощи класса BinarySerializer [3](#).
- класс Wrapper определяет бинарный объект, содержащий секции исходного файла (класс Section), которые в свою очередь состоят из полей разного типа (NestedField - для вложений, ValueField - поле ключ+значения, и CommentField - комментарий)

```

1 """ Конвертор
2 для обязательного задания, дополнительного задания №1 и №3. Без
3 использования сторонних библиотек осуществляется:
4 - десериализацию данных в формате INI,
5 - сериализацию в формат RON,
6 - сериализацию в формат XML
7 """
8
9 import os
10 from .convertor import Convertor
11 from typing import Sequence
12 from .formal import Wrapper, Section, ValueField, CommentField,
13     NestedField
14 from .formal import BinarySerializer
15
16 class HandWrittenConvertor(Convertor):
17     @staticmethod
18     def read_file(src: os.path):
19         with open(src, 'r', encoding='UTF-8') as f:
20             return list(filter(lambda row: row != "", map(lambda row: row.
21             strip(), f.readlines())))
22
23     @staticmethod
24     def deserialize(file_content: Sequence[str]) -> Wrapper:
25         def is_section(row):
26             return row.startswith('[') and row.endswith(']')
27
28         def is_nested(row):
29             prefix = row.split(',') [0]
30             return not is_section(row) and row.find(prefix) + len(prefix)
31             < row.find('=')
32
33         def find_subfields(prefix: str, fields: Sequence, nested_prefix:
34             str = "") -> NestedField:
35             if nested_prefix == "":
36                 nested_prefix = prefix
37                 nested = NestedField(key=nested_prefix)
38                 fields = list(filter(lambda f: f.split('=')[0].startswith(
39                     prefix), fields))
40                 for field in fields:
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136

```

```

37         full_field_key, value = field.split('=', )[:2]
38         actual_field_key = full_field_key[len(prefix) + 1:]
39
40         if '.' not in actual_field_key:
41             nested.add_field(ValueField(key=full_field_key.split(
42                 '.' )[-1], value=value))
43         else:
44             nested.add_field(find_subfields(full_field_key, fields,
45                 actual_field_key.split('.')[0]))
46     return nested
47
48     wrapper = Wrapper()
49
50     sections = '\n'.join(file_content).split(',')[1:]
51     for section_string in sections:
52         nested_fields = []
53         section_fields = list(filter(lambda r: r != '', section_string
54             .split('\n')))
55         section = Section(name=section_fields[0][: -1])
56
57         for i in range(1, len(section_fields)):
58             row = section_fields[i]
59             if i == len(section_fields) - 1:
60                 for nested in nested_fields:
61                     section.add_field(nested)
62                 nested_fields = []
63             wrapper.add_section(section)
64
65             if row[0] == '#' or row[0] == ';':
66                 section.add_field(CommentField(value=row[1:].strip()))
67             elif is_nested(row):
68                 if row.split('.')[0] in [field.get_key() for field in
69                     nested_fields]:
69                     continue
70                 nested_fields.append(find_subfields(prefix=row.split(
71                 '.' )[0], fields=section_fields))
72             else:
73                 section.add_field(ValueField(*row.split('=')))
74
75     return wrapper
76
77     @staticmethod
78     def serialize(obj: Wrapper, format: str):
79         format = format.lower()
80         if format == 'ron':
81             def write_field(out, key, value, type, indent):
82                 if type == "value_field":
83                     out.write(f'{indent}{key}: {value}\n')
84                     return
85                 if type == "comment_field":
86                     out.write(f'{indent}//{value}\n')
87                     return
88                 out.write(f'{indent}{key}: (\n')
89                 for field in value:
90                     write_field(out, field.get_key(), field.get_value(),
91                         field.get_type(), indent + "    ")
92                     out.write(f'{indent})\n')
93
94             output = os.path.abspath(os.path.join(os.path.dirname(
95                 __file__), 'output', 'schedule_hand_written.ron'))

```

```

90         with open(output, 'w', encoding='utf-8') as out:
91             out.write("Schedule(\n")
92             indent = "    "
93             for section in obj.get_sections():
94                 out.write(f"{indent}{section.get_name()}:\n")
95                 for field in section.get_fields():
96                     write_field(out, field.get_key(), field.get_value()
97 (), field.get_type(), indent + "    ")
98                     out.write(f"{indent}\n")
99             out.write(")\n")
100
101     elif format == 'xml':
102         def write_field(out, key, value, type, indent):
103             if type == "value_field":
104                 out.write(f"{indent}<{key}>{value}</{key}>\n")
105                 return
106             if type == "comment_field":
107                 out.write(f"{indent}<!-- {value} -->\n")
108                 return
109             out.write(f"{indent}<{key}>\n")
110             for field in value:
111                 write_field(out, field.get_key(), field.get_value(),
112 field.get_type(), indent + "    ")
113                 out.write(f"{indent}</{key}>\n")
114
115             output = os.path.abspath(os.path.join(os.path.dirname(
116 __file__), 'output', 'schedule_hand_written.xml'))
117             with open(output, 'w', encoding='utf-8') as out:
118                 out.write("""<?xml version="1.0" encoding="UTF-8"?>\n""")
119                 out.write("<main>\n")
120                 indent = "    "
121                 out.write(f"{indent}<Schedule>\n")
122                 for section in obj.get_sections():
123                     out.write(f"{indent*2}<{section.get_name()}>\n")
124                     for field in section.get_fields():
125                         write_field(out, field.get_key(), field.get_value()
126 (), field.get_type(), indent * 3)
127                         out.write(f"{indent*2}</{section.get_name()}>\n")
128                     out.write(f"{indent}</Schedule>\n")
129                     out.write("</main>\n")
130
131     else:
132         raise ValueError(f"Unknown format: {format}")

```

Листинг 2: hand_written_convertor.py

```

1 import pickle
2 from typing import List, Optional
3 from abc import ABC
4
5
6 class Field(ABC):
7     def __init__(self, key: str, value: str, type: str):
8         self.__key = key
9         self.__value = value
10        self.__type = type
11
12    def get_key(self) -> str:
13        return self.__key

```

```

15     def get_value(self) -> str:
16         return self.__value
17
18     def get_type(self) -> str:
19         return self.__type
20
21     def to_bytes(self) -> bytes:
22         # СерIALIZАЦИЯ объекта с помощью pickle
23         return pickle.dumps({
24             "key": self.__key,
25             "value": self.__value,
26             "type": self.__type,
27         })
28
29
30 class ValueField(Field):
31     def __init__(self, key: str, value: str):
32         super().__init__(key, value, "value_field")
33
34
35 class CommentField(Field):
36     def __init__(self, value: str):
37         super().__init__("!", value, "comment_field")
38
39
40 class NestedField(Field):
41     def __init__(self, key: str):
42         self.__value: List[Field] = []
43         super().__init__(key, self.__value, "nested_field")
44
45     def add_field(self, field: Field) -> None:
46         self.__value.append(field)
47
48     def get_fields(self) -> List[Field]:
49         return self.__value
50
51     def to_bytes(self) -> bytes:
52         # СерИализовать список вложенных полей
53         return pickle.dumps({
54             "key": self.get_key(),
55             "value": [f.to_bytes() for f in self.__value],
56             "type": self.get_type(),
57         })
58
59
60 class Section:
61     def __init__(self, name: str, fields: Optional[List[Field]] = None) ->
62         None:
63         self.__name = name
64         self.__fields = fields if fields is not None else []
65
66     def add_field(self, field: Field) -> None:
67         self.__fields.append(field)
68
69     def get_fields(self) -> List[Field]:
70         return self.__fields
71
72     def get_name(self) -> str:
73         return self.__name

```

```

74     def to_bytes(self) -> bytes:
75         # Сериализовать секцию: имя и список полей
76         return pickle.dumps({
77             "name": self.__name__,
78             "fields": [f.to_bytes() for f in self.__fields],
79         })
80
81
82 class Wrapper:
83     def __init__(self, sections: Optional[List[Section]] = None) -> None:
84         self.__sections = sections if sections is not None else []
85
86     def add_section(self, section: Section) -> None:
87         self.__sections.append(section)
88
89     def get_sections(self) -> List[Section]:
90         return self.__sections
91
92     def to_bytes(self) -> bytes:
93         return pickle.dumps(self)
94
95
96 class BinarySerializer:
97     @staticmethod
98     def save_to_file(obj: object, filename: str) -> None:
99         with open(filename, 'wb') as f:
100             f.write(obj.to_bytes())
101
102     @staticmethod
103     def load_from_file(filename: str) -> object:
104         with open(filename, 'rb') as f:
105             data = f.read()
106         return pickle.loads(data)

```

Листинг 3: formal.py

2. 2. Дополнительные задание 1

Разберем детально решение дополнительного задания 1.

- для конвертацию бинарного объекта, полученного в обязательном задании, в формат RON в классе HandWrittenConvertor был создан метод `serialize` [2](#)
- данный метод проходит в цикле по каждой секции бинарного объекта, определяет тип поля, и в зависимости от него добавляет соответствующий элемент RON

Результат: [..../src/output/schedule_hand_written.ron](#)

2. 3. Дополнительные задание 2

Разберем детально решение дополнительного задания 2.

- для сериализации и десериализации исходных данных был создан класс AdvancedConvertor в файле [4](#)
- данный класс содержит методы `read_file`, `deserialize` и `serialize` аналогичные HandWrittenConvertor, но написанные при помощи библиотек "configparser" для RON и "xml" для XML

Результат ручного перевода: [..../src/output/schedule_hand_written.ron](#)

Результат с использованием библиотек: [..../src/output/schedule_advenced.ron](#)

Результаты, полученные при помощи самописного парсера и парсера с использованием готовых библиотек имеют некоторые сходства и различия. Сходства состоят в том, что оба класса одинаково хорошо справляются с десериализацией исходного файла в бинарный объект.

Из различий можно выделить то, что самописный парсер добавляет необязательное имя структуры `Shedule` в начале RON файла ([1](#)), в то время как библиотека configparser этого не делает ([2](#)). Также самописный парсер позволяет обрабатывать вложенность в INI файле (разделение точкой), а библиотека configparser этого не делает.

Кроме того, самописный парсер умеет обрабатывать строки с комментариями, а configparser нет.

```

Schedule(
    Понедельник: {
        lesson1: (
            subject: Основы дискретной математики (базовый уровень)
            type: Лекции
            start: 09:50
            end: 11:20
            teacher: Поляков Владимир Иванович
            room: 2337
            building: Кронверкский пр., д.49, лит.А
        )
    }
)

```

Рис. 1: ручной перевод

```

[{
    Понедельник: {
        lesson1.subject: 'Основы дискретной математики (базовый уровень)',
        lesson1.type: 'Лекции',
        lesson1.start: '09:50',
        lesson1.end: '11:20',
        lesson1.teacher: 'Поляков Владимир Иванович',
        lesson1.room: '2337',
        lesson1.building: 'Кронверкский пр., д.49, лит.А',
    }
}]

```

Рис. 2: перевод через библиотеку

```

1 """ Конвертор
2 для дополнительного задания №2.С
3 использованием сторонних библиотек осуществляет:
4 - десериализацию данных в формате INI ,
5 - сериализацию в формат RON
"""
6
7
8 from .convertor import Convertor
9 import configparser
10 import xml.etree.ElementTree as ET
11 import os
12
13
14 class AdvancedConvertor(Convertor):
15     @staticmethod
16     def deserialize(parse_object: configparser.ConfigParser) -> dict:
17         return {section: dict(parse_object.items(section)) for section in
18                 parse_object.sections()}
19
20     @staticmethod
21     def serialize(object: dict, format: str, indent: int=0, return_value:
22         bool=False):
23         if format.lower() == 'ron':
24             ron = "{\n"
25             for key, value in object.items():
26                 if isinstance(value, dict):

```

```

26         ron += " " * (indent + 4) + f'{key}: {AdvancedConverter.
27     serialize(object=value, format='ron', indent=indent + 4, return_value=
28     True).strip()}\n"
29     else:
30         ron += " " * (indent + 4) + f'{key}: {repr(value)}},\n"
31     ron += " " * indent + "}"
32
33     if return_value:
34         return ron
35     output = os.path.abspath(os.path.join(os.path.dirname('__file__'), ',
36     output', 'schedule_advanced.ron'))
37     with open(output, "w", encoding='utf-8') as out:
38         out.write(ron)
39
40     elif format.lower() == 'xml':
41         root_name = 'main'
42         root = ET.Element(root_name)
43         for section, items in object.items():
44             section_elem = ET.SubElement(root, section)
45             for key, value in items.items():
46                 item_elem = ET.SubElement(section_elem, key)
47                 item_elem.text = value
48             if return_value:
49                 return root
50             output = os.path.abspath(os.path.join(os.path.dirname('__file__'), ',
51             output', 'schedule_advanced.xml'))
52             tree = ET.ElementTree(root)
53             tree.write(output, encoding="utf-8", xml_declaration=True)
54
55     @staticmethod
56     def read_file(src: os.path) -> configparser.ConfigParser:
57         object = configparser.ConfigParser()
58         object.read(src, encoding='utf-8')
59         return object

```

Листинг 4: advanced_convertor.py

2. 4. Дополнительные задание 3

Разберем детально решение дополнительного задания 3.

- для конвертацию бинарного объекта, полученного в обязательном задании, в формат XML в классе HandWrittenConvertor был изменен метод `serialize`
- теперь данный метод принимает аргумент "format" значение которого определяет перевод в нужный формат
- логика сериализации в XML практически не отличается от таковой в случае RON, за исключением видоизмененных полей файла в соответствии с форматом данных.

Результат: `../src/output/schedule_hand_written.ron`

2. 5. Дополнительные задание 4

Разберем детально решение дополнительного задания 4.

- для проведения тестов был создан класс TestRunner [5](#)
- данный класс содержит методы run_handwritten_test и run_advanced_test, которые осуществляют 100 итераций перевода при помощи рукописного парсера и парсера с использованием библиотек соответственно

Результаты 100 итераций таковы:

- для рукописного парсера - затрачено времени 88.102552 мс, 0.881025 мс на 1 итерацию
- для парсера с использованием библиотек - затрачено времени 120.904266 мс, 1.839042 мс на 1 итерацию

Таким образом, парсер с использованием библиотек выполняет перевод приблизительно на 36% медленнее. Это может быть связано с тем, что мой парсер сделан под узкую задачу, без дополнительной универсальной логики и проверок, характерных для библиотек, что уменьшает накладные расходы. Учитываются только нужные структуры и синтаксис, что ускоряет обработку и парсинг.

```
1 """ Тесты для дополнительного задания №4 """
2
3 import time
4
5 from lib.advanced_convertor import AdvancedConvertor
6 from lib.hand_written_convertor import HandWrittenConvertor
7 from lib.formal import BinarySerializer
8 from os import path
9
10
11 class TestRunner:
12     @staticmethod
13     def run_handwritten_test(src: path, bin_src: path, iter_num: int=100,):
14         start_time = time.time()
15
16         for _ in range(iter_num):
17             ini_file_content = HandWrittenConvertor.read_file(src)
18             serialized_raw = HandWrittenConvertor.serialize(file_content=
19             ini_file_content)
20             BinarySerializer.save_to_file(deserialized_raw, bin_src)
21             loaded_obj = BinarySerializer.load_from_file(bin_src)
22             HandWrittenConvertor.serialize(obj=loaded_obj, format='ron')
23
24         delta_time = (time.time() - start_time) * 1000
25         TestRunner.print_statistics(delta_time, "Собственный парсер")
26
```

```
27 @staticmethod
28 def run_advanced_test(src: path, iter_num: int=100):
29     start_time = time.time()
30
31     for _ in range(iter_num):
32         ini_parser = AdvancedConvertor.read_file(src)
33         serialized_advanced = AdvancedConvertor.serialize(parse_object=
34         ini_parser)
35         AdvancedConvertor.deserialize(serialized_advanced, format='ron')
36
37         delta_time = (time.time() - start_time) * 1000
38         TestRunner.print_statistics(delta_time, "С использованием библиотек")
39
40     @staticmethod
41     def print_statistics(delta_time, name):
42         print('='*40)
43         print(f"Результат 100 итераций - {name}")
44         print(f"Затрачено времени {delta_time} мс, {delta_time/100} мс на 1
итерацию")
```

Листинг 5: test.py

3. Вывод

В процессе выполнения лабораторной работы по информатике я ознакомился с формой Бэкуса-Наура, изучил основные принципы организации формальных грамматик [1], а также ознакомился и научился работать с незнакомыми мне ранее языками разметки,INI, XML и RON [2].

4. Литература

- [1] Орлов С. А., Цилькер Б. Я. Организация ЭВМ и систем: Учебник для вузов. 2-е изд. – СПб.: Питер, 2011. – 688 с.: ил., Приложение А «Арифметические основы вычислительных машин». URL: <https://bit.ly/4dzgo3u> (Дата обращения: 10.09.25)

- [2] Алексеев Е.Г., Богатырев С.Д. Информатика. Мультимедийный электронный учебник. Раздел 3 «Системы счисления». URL: <http://inf-e-alekseev.ru/text/Schisl.html> (Дата обращения: 10.09.25)