

4 Лабораторная работа №4. «Исследование протоколов, форматов обмена информацией и языков разметки документов»

4.1 Порядок выполнения работы

1. Определить номер варианта как остаток деления на 36 последних двух цифр своего идентификационного номера в ISU: например, 1255**98** / 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.
9. Дополнительное задание №1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
 - а) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
 - б) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.

- с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
10. Дополнительное задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
- а) Переписать исходный код, добавив в него использование регулярных выражений.
 - б) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
11. Дополнительное задание №3 (позволяет набрать +25 процентов от максимального числа баллов БаРС за данную лабораторную).
- а) 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.
 - б) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
 - с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
12. Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).
- а) Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.
 - б) Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
13. Дополнительное задание №5 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).
- а) 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.

- б) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

14. Проверить, что все пункты задания выполнены и выполнены верно.

15. Написать отчёт о проделанной работе.

16. Подготовиться к устным вопросам на защите.

4.2 Требования и состав отчёта

1. Отчёт должен быть выполнен на листе размером А4 с использованием Microsoft Word, Libre Office, Р7 Офис и т.п.
2. Цвет шрифта должен быть черным, размер шрифта – не менее 12 пт. Рекомендуемый тип шрифта для основного текста отчета – Times New Roman. Полужирный шрифт применяют только для заголовков разделов и подразделов, заголовков структурных элементов. Разрешается для написания определенных терминов, формул, теорем применять шрифты разной гарнитуры.
3. Текст отчёта следует печатать, соблюдая следующие размеры полей: левое – 30 мм, правое – 15 мм, верхнее и нижнее – 20 мм. Абзацный отступ должен быть одинаковым по всему документу и равен 1,25 см.
4. Отчёт должен содержать следующие разделы:
 - титульный лист;
 - содержание;
 - порядковый номер и название лабораторной работы;
 - задание (постановку задачи);
 - основные этапы вычисления (описание решаемой задачи, полный листинг программ .ру, содержание файла в исходном и результирующем форматах);
 - заключение;
 - список использованных источников.
5. Страницы работы нумеруются арабскими цифрами (нумерация сквозная по всему тексту). Номер страницы ставится в центре нижнего поля листа без точки. Титульный лист включается в общую нумерацию, номер на нем не ставится. Все страницы, начиная с 2-й, нумеруются.
6. Титульный лист должен быть оформлен в соответствии с следующей инструкцией:

- Весь текст должен быть выровнен по середине страницы.
 - В верхней части листа нужно написать полное наименование учебного заведения.
 - После названия вуза требуется написать название факультета (указывать название требуется, пропустив минимум одну строку после названия вуза).
 - В середине листа требуется указать вид работы (курсовой проект, компьютерный практикум или лабораторная работа).
 - Под видом работы указывается тема. Писать без кавычек.
 - Под темой указать номер варианта (при наличии).
 - Далее в правой части листа после слова «Выполнил» написать свои личные данные (ФИО, номер группы).
 - Под сведениями о студенте после слова «Проверил» указывается информация о преподавателе, который принимает работу. Помимо фамилии и инициалов преподавателя, указывается учёная степень (при наличии), учёное звание (при наличии) и должность.
 - В нижней части листа требуется указать город обучения и рядом через пробел написать год сдачи лабораторной работы.
7. Выравнивание текста по всей ширине листа.
 8. Отчет должен содержать автособираемое содержание (обязательные разделы – Задание, Основные этапы вычисления, Заключение, Список использованных источников).
 9. Отчет должен содержать рисунки, оформленные и подписанные в соответствии с ГОСТ 7.32-2017 «Отчет о научно-исследовательской работе. Структура и правила оформления» (**минимум один** рисунок): https://cs.msu.ru/sites/cmc/files/docs/2021-11gost_7.32-2017.pdf
 10. К рисункам относятся все графические изображения (схемы, графики, фотографии, рисунки, снимки экрана). На все рисунки в тексте должны быть даны ссылки. Рисунки должны располагаться непосредственно после текста, в котором они упоминаются впервые, или на следующей странице. Рисунки нумеруются арабскими цифрами, при этом нумерация сквозная, но допускается нумеровать и в пределах раздела (главы). В последнем случае номер рисунка состоит из номера главы и порядкового номера иллюстрации, разделенных точкой (например: Рисунок 1.1). Название пишется под

рисунком по центру, как и рисунок, форматирование — как и у обычного текста. Слово «Рисунок» пишется полностью. Если рисунок один, то он обозначается «Рисунок 1». В тексте отчёта должны быть перекрёстные ссылки на рисунок/рисунки.

11. Отчет должен содержать таблицы, оформленные и подписанные в соответствии с ГОСТ 7.32-2017 «Отчет о научно-исследовательской работе. Структура и правила оформления» (**минимум одна** таблица): https://cs.msu.ru/sites/cmc/files/docs/2021-11gost_7.32-2017.pdf

12. Отчет должен содержать список литературы со ссылками на источники (**минимум два** источника). В тексте отчёта должны быть перекрёстные ссылки на список источников. Список использованных источников оформлять согласно следующим документам:

- ГОСТ 7.1 — 2003 «Библиографическая запись. Библиографическое описание. Общие требования и правила составления» <https://www.csu.ru/faculties/Documents/gost%207.1-2003.pdf>
- ГОСТ 7.12 — 77 «Сокращение русских слов и словосочетаний в библиографическом описании» <https://files.stroyinf.ru/Data2/1/4294822/4294822775.pdf>
- ГОСТ 7.11—78 «Сокращение слов и словосочетаний на иностранных языках в библиографическом описании» <https://files.stroyinf.ru/Data2/1/4294822/4294822776.pdf>
- ГОСТ 7.80 — 2000 «Библиографическая запись. Заголовок. Общие требования и правила составления» http://diss.rsl.ru/datadocs/doc_291ku.pdf
- ГОСТ 7.82 — 2001 «Библиографическое описание электронных ресурсов: общие требования и правила составления» <http://www.lib.tsu.ru/win/metod/gost/gost7.82-2001.pdf>
- ГОСТ 7.0.5 — 2008 «Библиографическая ссылка. Общие требования и правила составления» http://diss.rsl.ru/datadocs/doc_291tu.pdf

13. Отчёт предоставить в электронном виде (записать на флэш-накопитель и продублировать себе на электронную почту/файлообменник/облачное хранилище). По требованию преподавателя нужно быть готовыми скомпилировать и запустить свою программу на компьютере в учебной аудитории (или своём ноутбуке).

4.3 Подготовка к защите

1. Изучить и закрепить необходимый материал из следующего пособия: Лямин А.В., Череповская Е.Н. Объектно-ориентированное программирование. Компьютерный практикум. – СПб: Университет ИТМО, 2017. – 143 с. – Режим доступа: <https://books.ifmo.ru/file/pdf/2256.pdf>.
2. Прочитать и повторить информацию из статьи в Википедии: https://ru.wikipedia.org/wiki/Форма_Бэкуса_—_Наура.
3. Прочитать и повторить информацию из статьи в Википедии: https://ru.wikipedia.org/wiki/Формальная_грамматика.
4. Прочитать и повторить информацию из статьи «Пишем изящный парсер на Питоне»: <https://habr.com/ru/post/309242/>.
5. Уметь объяснить каждую строку программы, представленной в отчёте.
6. При защите отчёта надо уметь отвечать на вопросы по работе программы, вопросы по материалам лекций №3 и №4 и следующие вопросы:
 - 1) В чём разница между Markup и Markdown?
 - 2) В чём заключается особенность PROTOBUF по сравнению с другими форматами?
 - 3) Чем формат CSV отличается от формата TSV?
 - 4) Чем обусловлено постоянное появление новых форматов представления данных?
 - 5) Каким образом в формате XML представляются символы '>' и '<'?
 - 6) Что такое сериализация данных?
 - 7) Каким образом в YAML обозначаются комментарии?
 - 8) Пояснить, как в языке разметки Markdown создать заголовки разных уровней, оформить код, вывести полужирный, курсивный и зачёркнутый текст?
 - 9) Какие форматы обмена данных используются в современных популярных мессенджерах (Viber, WhatsApp, Telegram и т.д.)?
 - 10) Как расшифровывается аббревиатура SVG?
 - 11) Привести пример использования в языке HTML тега, который создаёт гиперссылку на url.
 - 12) Какое число структур может представлять собой в закодированном виде JSON-текст?

4.4 Варианты заданий

№ варианта	Исходный формат	Результирующий формат	Дни недели
0	JSON	XML	Понедельник, четверг
1	XML	JSON	Понедельник, четверг
2	JSON	YAML	Понедельник, четверг
3	YAML	JSON	Понедельник, четверг
4	YAML	XML	Понедельник, четверг
5	XML	YAML	Понедельник, четверг
6	JSON	XML	Вторник, пятница
7	XML	JSON	Вторник, пятница
8	JSON	YAML	Вторник, пятница
9	YAML	JSON	Вторник, пятница
10	YAML	XML	Вторник, пятница
11	XML	YAML	Вторник, пятница
12	JSON	XML	Среда, суббота
13	XML	JSON	Среда, суббота
14	JSON	YAML	Среда, суббота
15	YAML	JSON	Среда, суббота
16	YAML	XML	Среда, суббота
17	XML	YAML	Среда, суббота
18	JSON	XML	Вторник, четверг
19	XML	JSON	Вторник, четверг
20	JSON	YAML	Вторник, четверг
21	YAML	JSON	Вторник, четверг
22	YAML	XML	Вторник, четверг
23	XML	YAML	Вторник, четверг
24	JSON	XML	Среда, пятница
25	XML	JSON	Среда, пятница
26	JSON	YAML	Среда, пятница
27	YAML	JSON	Среда, пятница
28	YAML	XML	Среда, пятница
29	XML	YAML	Среда, пятница

30	JSON	XML	Понедельник, суббота
31	XML	JSON	Понедельник, суббота
32	JSON	YAML	Понедельник, суббота
33	YAML	JSON	Понедельник, суббота
34	YAML	XML	Понедельник, суббота
35	XML	YAML	Понедельник, суббота