Теория Формальных Языков.

Лабораторная работа №2.

Масягин М. М.

ИУ9-52Б

Задание-1. RegExp почтовых кодов.

Так как в задании не сказано, какие именно почтовые коды нужно искать, то я решил искать Российские. Российские почтовые коды — это наборы из 6 цифр, причем имперически было установлено, что первая цифра бывает лишь [1-46]. Поэтому регулярное выражение на Python для выделения выглядит следующим образом:

**r'\b[1-46]\d{5}\b'**, где:

**\b** — граница слова (если не использовать \b, то 228148832282 будет содержать 2 валидных почтовых кода, хотя это просто 12-значное число). Python убирает границы слова, то есть если мы нашли по Regexp`у строку « 228322 », то нам вернется «228322»;

**[1-46]** — перечисление цифр 1,2,3,4,6;

**\d{5}** — строгое повторение 5 раз любых цифр от 0 до 9;

Зададим регулярное выражение формально:

**\b[1-46][0-9][0-9][0-9][0-9][0-9]\b**

в классических регулярных выражениях (итерация, конкатенация, объединение) нет Python`овского **\b**, поэтому нам нужно отбросить первый и последний символ.

Контрольный пример — просто куча слов, с вкраплениями различных 6-значных чисел, находится в файле postcodes/test/test.txt.

Код программы лежит в файле postcodes/regexp.py.

Задание-2. Лексер.

Мне достался лексер языка JavaScript. При написании, я использовал библиотеку ply — биндинг для GNU FLEX и GNU YACC. Таким образом, мне не требуется переводить программу в «набор токенов», мне достаточно лишь упорядоченной по порядку следования таблицы токенов.

Лексер отбрасывает комментарии;

В комплект моей библиотеки-лексера входит функция для нормализации пробелов, символов новой строки, табуляций и тд;

Лексер способен находить ошибки в виде «необрабатываемых» символов, например «@»;

Формат таблицы лексем стандартен для GNU FLEX ply:

LexToken(VAR,'var',1,0)

LexToken(ID,'x',1,4)

LexToken(EQ,'=',1,6)

LexToken(NUMBER,'5',1,8)

LexToken(SEMI,';',1,9) для кода var x = 5;

Каждый LexToken содержит:

1) Тип токена.

2) Какая последовательность символов программы ему соответствует.

3) Номер строки.

4) Номер столбца.