Пояснение к практикуму.

Михайлов Никита 696

27 октября 2017 г.

Пусть есть какое-то регулярное выражение и какое-то слово. Далее пусть n - длина слова word. Построим матрицу для конкретного регулярноего выражения и слова размера $(n+1) \times n$, где $a[i][j] = +$, если с помощью регулярного выражения можно вывести j символов слова начиная с позиции i. $a[i][j] = -$ иначе.

Будем находить ответ по индукции по длине регулярного выражения, поддерживая инвариант, описанный выше. Будем пересчитывать матрицы для регулярных подвыражений, которые возникают в процессе обработки регулярного выражения, поступившего на вход. В конце получим матрицу, соотвествующую всему регулярному выражению.

База:

Так как в алфавите только буквы a, b, c, то посчитаем матрицы с вышеописанным инвариантом и сохраним их (чтобы не пересчитывать их в дальнейшем). Для пустого слова в матрице в первой строке (у строки индекс 0) проставим все $+$.

Переход:

Рассмотрим как пересчитывать матрицы для таких функций как конкатенация, $+$, *.

1) Для конкатенации. Пусть есть два выражения A и B со своими матрицами. Пусть в матрице A стоит $+$ на позиции i, j. Это означает, что начиная с позиции i может вывести j символов. Подумаем, сколько еще букв мы можем вывести с помощью регулярного выражения B. Для этого посмотрим в матрице B какое максимальное количество букв можно вывести с позиции i + j.

2) Для $+$ делаем глобальный OR для матриц A и B.

3) Для * заметим что с любой позиции выводится пустое слово. Далее поступим так же как и с конкатенцией, только вместо матрицы B будем использовать ту же самую матрицу.

Инвариант сохраняется из постоения.

Если слово на входе не равно пустому слову, то понятно, что слово принадлежит языку тогда и только тогда, когда в таблице в последней строчке в первой столбце стоит $+$. Если слово на входе равно пустому слову, то стоит смотреть в первую строчку матрицы в первый столбец. Само пустое слово будем обозначать как 1.