Практикум 1

Кушнерюк Сергей, Б05-925

1 Problem

Была предложена следующая задача:

Пусть у нас есть регулярное выражение в польской нотации, которое определяет язык L и некоторое слово W. Как найти максимальное неотрицательное целое число K такое, что W^K - это некоторое подслово $U \in L$

2 Solution

В силу эквивалентности классов регулярных и автоматных языков, предположим, что мы уже имеем автомат $\langle Q, \Sigma, \Delta, q_0, F\rangle$, задающий тот же язык, что и регулярное выражение, причем в нем нет бесполезных состояний v, т.е. таких что $\nexists q_F \in F \nexists U \in \Sigma^* : <v, U> \vdash <q_F, \varepsilon>$ или $\nexists U \in \Sigma^* : <q_0, U> \vdash <v, \varepsilon>$

Предположим, что мы знаем, для каждого состояния в какие состояния мы можем прийти по слову W. Рассмотрим граф $G : V = Q$, т.е. с вершинами соответствующими состояниям автомата. Ориентированное ребро $v \to u$ в графе будет проведено тогда и только тогда, когда $<v, W> \vdash <u, \varepsilon>$. Тогда путь наибольшей длины в таком графе будет соответствовать слову содержащее W^K, где K максимально. Действительно, рассмотрим первое и последнее состояние в пути.

Т.к. нет бесполезных состояний, то существует слово, по которому можно добраться до первого состояния, и слово, по которому можно добраться до какого-то из завершающих состояний. Тогда в силу построения графа существует слово, содержащее W^K. Предположим, что есть слово, которое содержит большую степень W^M. Тогда найдутся состояния, по которым возможно получить слово в большей степени, а значит в силу построения графа G длина наибольшего пути в нем не может быть больше K. Также, если в графе окажется цикл, то это значит, что максимума не существует, т.е. возможно получить слово, содержащее W в произвольной степени.

Таким образом, поиск наибольшего такого K сводится к нахождению наибольшего пути в ориентированном графе и проверку на наличие в нем цикла. Проверку на наличие цикла проведем с помощью поиска в глубину [1]. Также, с помощью него же построим топологическую сортировку вершин графа и найдем в порядке топологической сортировки путь максимальной длины в графе [2]

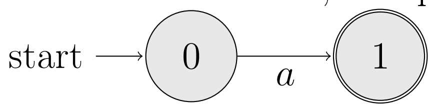
Осталось только построить автомат по польской нотации. Будем строить рекурсивно.

Если есть два автомата, построенных по польской нотации, конкатенация двух автоматов - это проведение всевозможных ребер по пустому слову из завершающих состояний первого автомата в стартовое второго.

Если есть два автомата, построенных по польской нотации, сумма двух автоматов - это добавление нового стартового состояния, из которого проведенены ребра по пустому слову в стартовые состяния исходных автоматов, а также объединение (в смысле множеств) завершающих состояний

Если есть автомат, то его итерация - это добавление нового стартового состояния, из которого проведенено ребро по пустому слову в стартовые состяния исходных автоматов, а также ребра по пустому слову из всех завершающих исходного слова в новое стартовое

База: это автоматы, которые дают единственное слово из одной буквы



(Асимптотики в процессе)

3 References

1. Проверка графа на ацикличность и нахождение цикла

2. Longest path problem. Acyclic graphs and critical paths