Шаныгин Иван Б09
Задание 1(*) из ДЗ №4 Доказательства нерегулярности языков. Минимизация.

Реализуйте алгоритм Хопкрофта, сравните время его работы с реализованным алгоритмом Мура.

Докажем, что время работы алгоритма Хопкрофта составляет $O(n \log n)$

Лемма 1.

Количество классов, созданных во время выполнения алгоритма, не превышает $2|Q| - 1$.

Доказательство: рассмотрим дерево, которое соответствует операциям разделения классов на подклассы. Корнем этого дерева является все множество состояний Q. Листьями являются классы эквивалентности, оставшиеся после работы алгоритма. Так как дерево бинарное — каждый класс может иметь лишь два потомка, а количество листьев не может быть больше $|Q|$, то количество узлов этого дерева не может быть больше $2|Q| - 1$.

Лемма 2.

Количество итераций цикла *while* не превышает $2|\Sigma||Q|$.

Доказательство: Для доказательства этого утверждения достаточно показать, что количество пар $\langle C, a\rangle$ добавленных в очередь *Queue* не превосходит $2|\Sigma||Q|$, так как на каждой итерации мы извлекаем одну пару из очереди. По лемме 1 количество классов не превосходит $2|Q| - 1$. Пусть C элемент текущего разбиения. Тогда количество пар $\langle C, a\rangle$, $a \in \Sigma$ не может быть больше $|\Sigma|$. Отсюда следует, что всего различных пар, которые можно добавить в очередь, не превосходит $2|\Sigma||Q|$.

Лемма 3.

Пусть $a \in \Sigma$ и $p \in Q$. Тогда количество пар $\langle C, a\rangle$, где $p \in C$, которые мы удалим из очереди, не превосходит $\log_2(|Q|)$ для фиксированных a и p.

Доказательство: Рассмотрим пару $\langle C, a\rangle$, где $p \in C$, которую мы удаляем из очереди. И пусть $\langle C', a\rangle$ следующая пара, где $p \in C'$ и которую мы удалим из очереди. После первого же разбиения класса C на подклассы мы добавим в очередь пару $\langle C'', a\rangle$, где C'' меньший из образовавшихся подклассов, то есть $|C''| \leq |C|/2$. Так же заметим, что $C' \subset C''$, а следовательно $|C'| \leq |C|/2$. Но тогда таких пар не может быть больше, чем $log_2(|Q|)$.

Лемма 4.

Σ |Inv| по всем итерациям цикла *while* не превосходит $|\Sigma||Q| \log_2(|Q|)$.

$\text{Inv} = \{q \mid q \in Q, \delta(q, a) \in C\}$
Доказательство: Пусть $x, y \in Q$, $a \in \Sigma$ и $\delta(x, a) = y$. Зафиксируем эту тройку. Заметим, что количество раз, которое x встречается в Inv при условии, что $\delta(x, a) = y$, совпадает с числом удаленных из очереди пар $\langle C, a\rangle$, где $y \in C$. Но по лемме 3 эта величина не превосходит $\log_2(|Q|)$. Просуммировав по всем $x \in Q$ и по всем $a \in \Sigma$ мы получим утверждение леммы.

Теорема.

Время работы алгоритма Хопкрофта равно $O(|\Sigma||Q| \log(|Q|))$.

Доказательство: Оценим, сколько времени занимает каждая часть алгоритма: Построение массива Inv занимает $O(|\Sigma||Q|)$ времени. По второй лемме количество итераций цикла *while* не превосходит $O(|\Sigma||Q|)$. Операции с множеством T' и разбиение классов на подклассы требуют $O(\Sigma(|Inv|))$ времени. Но по лемме 4 $\Sigma(|Inv|)$ не превосходит $|\Sigma||Q| \log_2(|Q|)$, то есть данная часть алгоритма выполняется за $O(|\Sigma||Q|log_2(|Q|))$. В лемме 1 мы показали, что в процессе работы алгоритма не может появится больше, чем $2|Q| - 1$ классов. Итого, получается, что время работы алгоритма Хопкрофта не превышает $O(\Sigma|Q|) + O(\Sigma|Q|) + O(\Sigma|Q| \log_2(|Q|)) + O(|\Sigma|Q|) = O(|\Sigma||Q| \log_2(|Q|))$.

Из лекций мы знаем, что алгоритм Мура работает за $O(|\Sigma| n^2)$. Алгоритм Хопкрофта работает быстрее - $O(|\Sigma| n \log n)$ (Однако существует алгоритм Бржозовского, который работает быстрее Хопкрофта при определенных параметрах автомата). Алгоритм Хопкрофта точно так же,как и Мур, последовательно расщепляет классы эквивалентности, но за счёт хитроумных структур данных он меньше тратит времени на просмотр классов, которые расщепить не получается или пока не получается.