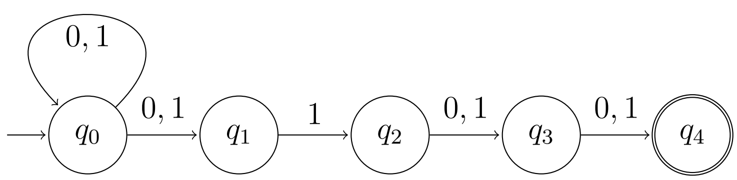
Задание 3

НКА и ДКА
Лемма о накачке

Автоматы и распознавание текстов
Усвяцов Михаил Рудольфович, 176 б

Задача 1. Определим язык $L_i = \{w \mid |w| = n, w[n-i] = 1\}$, то есть в язык L_i входят все слова, в которых 1 стоит на i-ом месте от конца[1]. Постройте НКА, распознающий язык L_3. По построенному НКА постройте ДКА.

Построение НКА для заданного языка не вызывает трудностей.



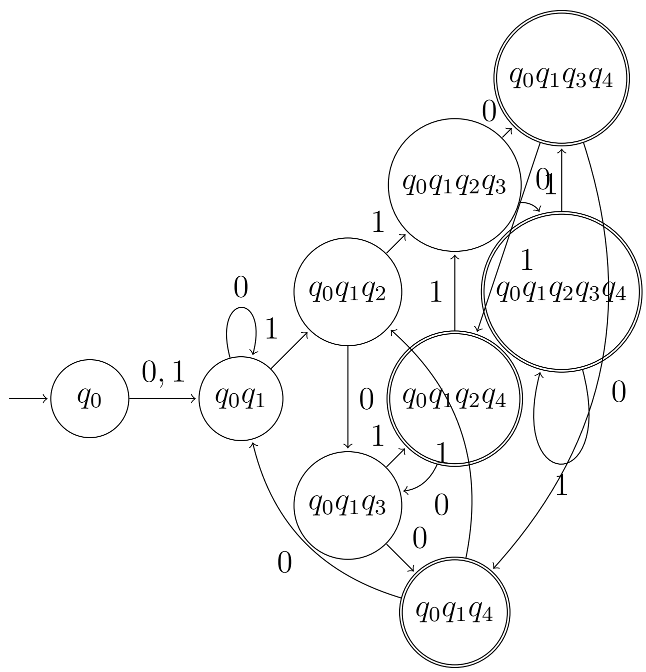
Для построения ДКА будем придерживаться алгоритма из книги Серебрякова.

1. Так как из q_0 нет ε переходов, то помечаем q_0 как непомеченное.
2. $\text{move}(q_0,\, 0) = \{q_0, q_1\}$.
3. Ставим множество $\{q_0, q_1\}$ в очередь непемеченных.
4. $\text{move}(q_0,\, 1) = \{q_0, q_1\}$.
5. $\text{move}((q_0, q_1),\, 0) = \{q_0, q_1\}$.
6. $\text{move}((q_0, q_1),\, 1) = \{q_0, q_1, q_2\}$.
7. Ставим множество $\{q_0, q_1, q_2\}$ в очередь непемеченных.

[1]Во избежании путаницы, первый с конца символ – это последний символ слова.

8. $\text{move}((q_0, q_1, q_2),\ 0) = \{q_0, q_1, q_3\}$.
9. Ставим множество $\{q_0, q_1, q_3\}$ в очередь непемеченных.
10. $\text{move}((q_0, q_1, q_2),\ 1) = \{q_0, q_1, q_2, q_3\}$.
11. Ставим множество $\{q_0, q_1, q_2, q_3\}$ в очередь непемеченных.
12. $\text{move}((q_0, q_1, q_3),\ 0) = \{q_0, q_1, q_4\}$.
13. Ставим множество $\{q_0, q_1, q_4\}$ в очередь непемеченных.
14. $\text{move}((q_0, q_1, q_3),\ 1) = \{q_0, q_1, q_2, q_4\}$.
15. Ставим множество $\{q_0, q_1, q_2, q_4\}$ в очередь непемеченных.
16. $\text{move}(q_0, q_1, q_2, q_3,\ 0) = \{q_0, q_1, q_3, q_4\}$.
17. Ставим множество $\{q_0, q_1, q_3, q_4\}$ в очередь непемеченных.
18. $\text{move}((q_0, q_1, q_2, q_3),\ 1) = \{q_0, q_1, q_2, q_3, q_4\}$.
19. Ставим множество $\{q_0, q_1, q_2, q_3, q_4\}$ в очередь непемеченных.
20. $\text{move}((q_0, q_1, q_4),\ 0) = \{q_0, q_1\}$.
21. $\text{move}((q_0, q_1, q_4),\ 1) = \{q_0, q_1, q_2\}$.
22. $\text{move}((q_0, q_1, q_2, q_4),\ 0) = \{q_0, q_1, q_3\}$.
23. $\text{move}((q_0, q_1, q_2, q_4),\ 1) = \{q_0, q_1, q_2, q_3\}$.
24. $\text{move}((q_0, q_1, q_3, q_4),\ 0) = \{q_0, q_1, q_4\}$.
25. $\text{move}((q_0, q_1, q_3, q_4),\ 1) = \{q_0, q_1, q_2, q_4\}$.
26. $\text{move}((q_0, q_1, q_2, q_3, q_4),\ 0) = \{q_0, q_1, q_3, q_4\}$.
27. $\text{move}((q_0, q_1, q_2, q_3, q_4),\ 1) = \{q_0, q_1, q_2, q_3, q_4\}$.

Теперь, когда мы описали все функции переходов, изобразим полученный ДКА.



Задача 2*. Докажите, что на языках L_i между НКА и построенными по ним ДКА достигается экспоненциальный разрыв.

Исходя из постановки задачи, языку принадлежат все слова, у которых на месте i с конца стоит 1. При этом значения в остальных позициях значения не имеют. Тогда понятно, что ДКА, эквивалентный этому языку должен помнить все последние i символов. Всего имеется 2^i последовательностей, состоящих из из последних i символов. Таким образом, Если в ДКА меньше 2^i состояний, то нашлось бы такое состояние q, что автомат попадает в него при прочтении двух разных последовательностей a и b. Рассмотрим такие последовательности a и b, что $a_i \neq b_i$. Теперь очевидно, что какое бы мы не задали i, автомат обязан иметь состояние, которое одновременно будет приемным и неприниемным, так как через одино и то же состояние q будет проходить цепочка удовлетворяющее языку и цепочка ему неудовлетворяющие. Из этого противоречия можно заключить, что число состояний ДКА должно быть 2^i. ЧТД.

Задача 3. Будут ли регулярными следующие языки?

1. $L = \{a^{2013n+5} \mid n = 0, 1, \} \cap \{a^{503k+29} \mid k = 401, 402, \ldots\} \subseteq \{a^*\}$.

Да, регулярен. Я нашел общее решение диафантового уравнения $a^{2013n+5} = a^{503k+29}$. При заданных n и k решений этого уравнения не существует, следовательно, пересечение первых двух множеств пусто. Пустое множество регулярно по определению регулярного языка.

2. $L_2 = \{a^{200n^2+1} \mid n = 1000, 1001, \ldots\} \subseteq \{a^*\}$.

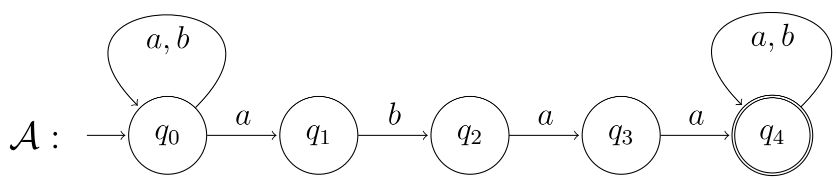
Применим лемму о накачке.

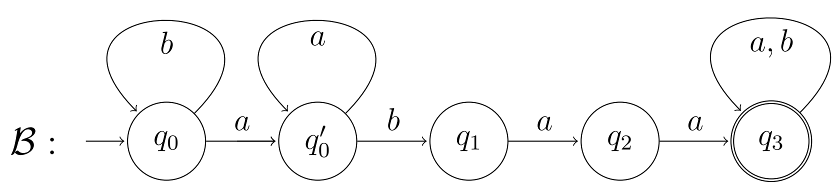
Разобьем наш язык на xyz так, что $y = a^k, k \geqslant 1$. Тогда любое слово нашего языка можно будет записать как: $a^{200n^2+1-k-m}a^ka^m$. А теперь прокачем y: $a^{200n^2+1-k-m}a^{2k}a^m$ уже не принадлежит нашему языку. Значит, язык нерегулярен.

3. Язык L_3 всех слов в алфавите $\{0, 1\}$, которые представляют числа в двоичной записи, дающие остаток два при делении на три (слово читается со старших разрядов). Например, 001010 ($1010_2 = 10_{10} = 3\times3+1$) $\notin L_3$, а 10001 ($10001_2 = 17_{10} = 5 \times 3 + 2$) $\in L_3$.

Воспользуемся отрицанием леммы о накачке. Рассмотрим слово 101. Пусть y = 0. Но 1001 уже не принадлежит языку L_3. Следовательно, язык нерегулярен.

Задача 4. Постройте по НКА $\mathcal{A}$ ДКА $\mathcal{B}$.



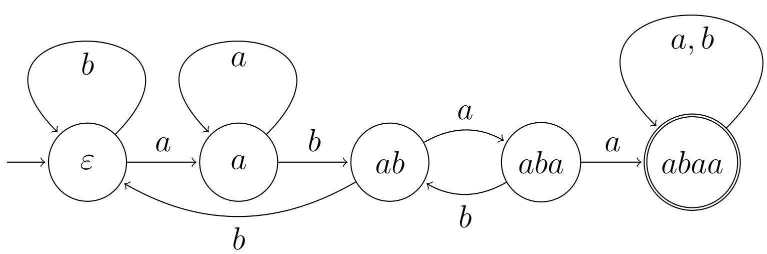


Задача 5*. Докажите, что в ДКА, распознающем язык $\Sigma^*w\Sigma^*$ не может быть меньше состояний чем элементов последовательности $l(w), l(l(w)), \ldots$

Для начала выбросим из рассмотрения Σ^*, так как мы рассматриваем минимальное возможное количество состояний. Теперь осталось доказать, что ДКА, который распознает w не может иметь меньше состояний, чем элементов в последовательности $l(w), l(l(w)), \ldots$.

Рассмотрим строку w = a. Для разбора такой строки необходимо 2 состояние ДКА. l(a) = ε. Следовательно, условия задачи выполняются. Теперь предположим, что для строки $w \in a|b^*$ длины n условия задачи выполняется, то есть для разбора строки w ДКА необходимо больше или равно состояний, чем в последовательности собственных префиксов слова w. Теперь рассмотрим случай с добавлением нового сивола в конец или начало слова. Приписывание справа буквы a могло увеличить количество префиксов. Тогда рассмотрим, сколько максимально могло бы быть собственных префиксов у нового слова, исходя из его длины. $(a^n) = a^{n-1}$. Следовательно, для слова длиной n + 1 максимально возможное число элементов в последовательности собственных префиксов не превышает n + 1. ДКА для разбора нового слова так же должен иметь n + 2 состояний. ЧТД.

Задача 6. Постройте КМП-автомат для слова *abaa*.



Задача 7*. Докажите, что КМП-автомат для слова w распознаёт язык $\Sigma^* w \Sigma^*$.

1 Дополнительные задачи

Задача 8. Приведите протокол работы КМП-алгоритма при поиске подслова *abba* в слове *abbbababbab*.

Начнем разбирать строчку $abbaTabbbababbab$

1. l[1] = l("a").length = 0
2. l[2] = l("ab").length = 0
3. l[3] = l("abb").length = 0
4. l[4] = l("abba").length = 0
5. l[5] = l("abbaT").length = 0
6. l[6] = l("abbaTa").length = 1
7. l[7] = l("abbaTab").length = 2
8. l[8] = l("abbaTabb").length = 3
9. l[9] = l("abbaTabbb").length = 0
10. l[10] = l("abbaTabbba").length = 1
11. l[11] = l("abbaTabbbab").length = 2
12. l[12] = l("abbaTabbbaba").length = 1
13. l[13] = l("abbaTabbbabab").length = 2
14. l[14] = l("abbaTabbbababb").length = 3
15. l[15] = l("abbaTabbbababba").length = 4
16. l[16] = l("abbaTabbbababbab").length = 2
17. Мы можем сказать, что подтсрока входит в строку тогда и только тогда, когда в массиве l есть число равное длине подстроки. Длина подстроки - 4. Следовательно, она найдена