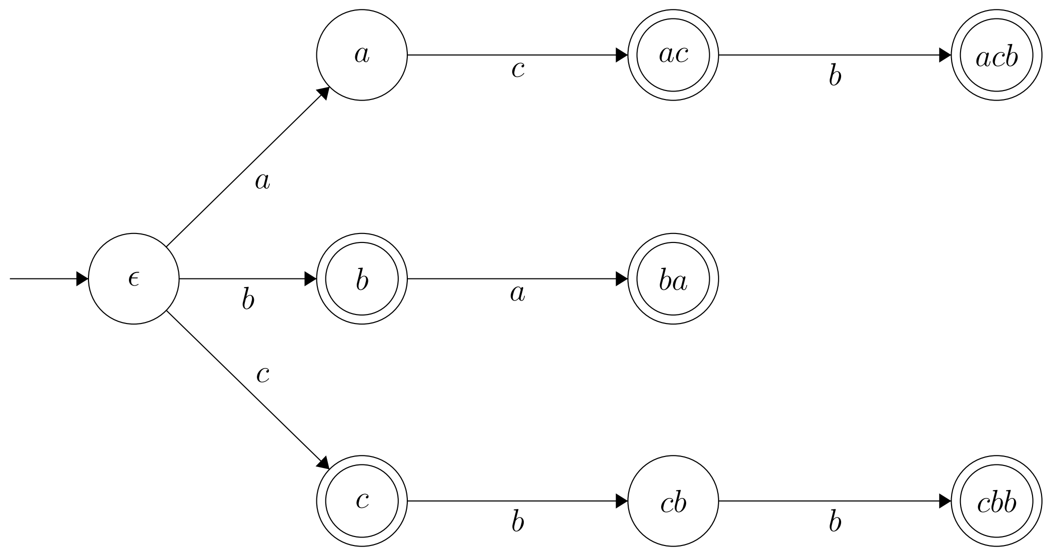
Домашняя работа по ТРЯП №4

Автор - Айвазов Денис из 671 группы

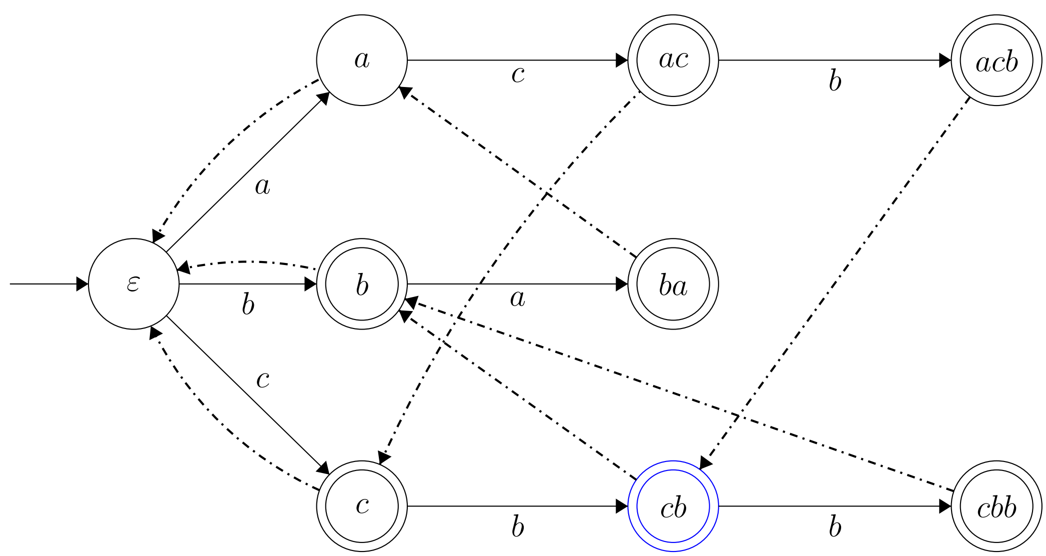
27 сентября 2017

1 НКА по языку L_3. Построить ДКА по НКА

Автомат для словаря $S = \{ac, acb, b, ba, c, cbb\}$был таким:

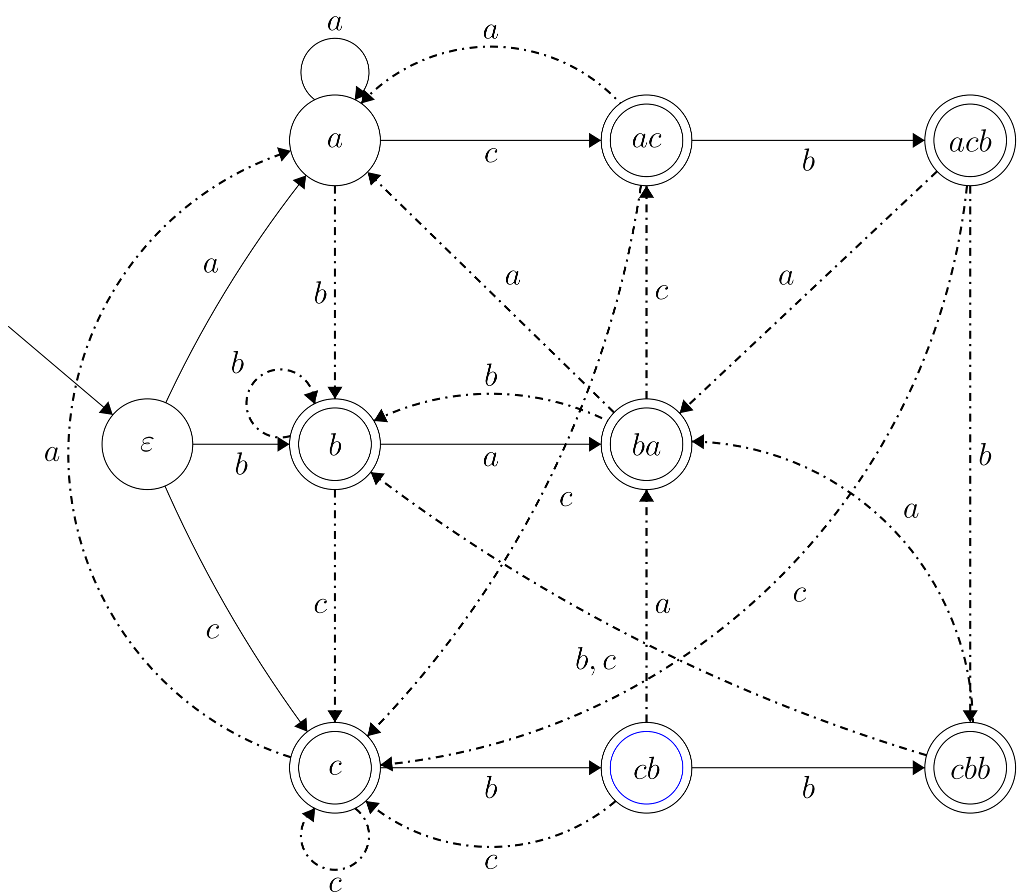


По алгоритму с листочка построим автомат Ахо-Корасик.



Вот протокол работы автомата Ахо Корасик на слове $acbacbb$:
$\varepsilon(a) \Rightarrow a(c) \Rightarrow \mathbf{AC}(\mathbf{b}) \Rightarrow \mathbf{ACB}() \dashrightarrow \mathbf{cb}() \dashrightarrow \mathbf{B}(\mathbf{A}) \Rightarrow \mathbf{BA}() \dashrightarrow a(c) \Rightarrow \mathbf{AC}(\mathbf{b}) \Rightarrow \mathbf{ACB}() \dashrightarrow \mathbf{cb}() \Rightarrow \mathbf{CBB}()$

Заглавными буквами отмечены изначально(у старого автомата) принимающие состояния. Проходясь по ним, с некоторой оговоркой мы получим, что ac,acb и были в слове 2 раза, b - три раза, а ba и cbb - 1. А оговорка заключается в следующем. При каждой суффиксной ссылке надо считать, что мы дошли из эпсилон до данного состояния. И учли все принимающие состояния. Ведь при обработке acb мы пропускаем состояние c, потому что буква c является подсловом слов ac, acb и cb. Но если каждый раз отсчитывать наш путь с начала, то мы посчитаем все вхождения подслов корректным образом.
Заметим, что cb тоже стало принимающим состоянием. А вот ДКА Ахо-Корасик, по алгоритму с листочка.



2 Проверить на регулярность

2.1 $L_1 = \{a^{2017n+5} | n = 0, 1, \dots\} \cap \{a^{503k+29} | k = 401, 402, \dots\};$

Сначала докажем, что пересечение не пусто (т.е. что сам язык не пустой).
Это можно сделать по расширенному Алгоритму Евклида, рассказанному на курсе Алгоритмы Построение и Анализ во 2ом семестре.
$a = 2017; b = -503; c = 24;$ решая Диофантово уравнение по алгоритму получим, что решения имеют вид:
$x = -4824 - 503m;\ y = -19344 - 2017m;\ \forall m \in \mathbb{Z}$
Или, для условий нашей задачи:
$x_0 = 206 + 503q \geq 0;\ y_0 = 826 + 2017q \geq 401;\ \forall q \in \mathbb{N}_0$ Т.е. $2017x_0 + 5 = 503y_0 + 29;$ - а это и степени буквы а в словах языка. Тогда наш язык состоит из слов вида, при заданных n и k, являющихся решениями Диофантова уравнения :

$$a^{2017n+5} = a^{503k+29} = a^{2017\cdot(206+q503)} = a^{2017\cdot 206}(a^{503\cdot 2017})^*$$

А это и есть РВ, задающее наш язык. А по теореме, доказанной на семинаре и в книге Хопкрофт, Мотвани, Ульман, РВ это способ задания регулярного языка. Значит мы доказали, что язык $L_1 \in REG$, поскольку он может быть задан РВ.

2.2 $L_2 = \{a^{200n^2+1} | n = 1000, 1001, \dots\} \subset a^*$

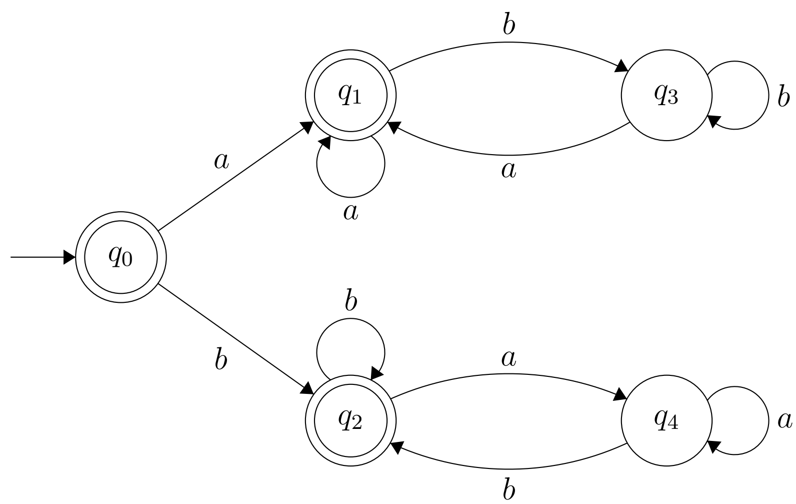
Докажем нерегулярность от противного: Если допустить, что $L_2 \in REG$, то по лемме о накачке сразу получии, что $\exists p\ \ \forall \omega \in L_2 :\ |\omega| > p$ выполнено $\omega = xyz;\ \ |y| > 0;\ |xy| \leq p;\ \ \forall i \geq 0\ \ xy^iz \in L_2$.
Из описания языка сразу видно, что все слова имеют вид $\omega = a^q$, где $q = 200n^2 + 1;\ \ n \in \mathbb{N}; n \geq 1000;$ Тогда и $y = a^q$
Тогда любое фикс $\omega = xy^iz = a^{l+iq}$ тоже должно принадлежать L_2. Но из-за квадратичного роста длины слова, оно перестанет влезать. т.е. $\omega_k - \omega_{k-1} \geq m$. т.е. разница между соседними словами поочереди начиная с какого-то k будет переваливать за m. А по своей сути это значит, что это слово уже не будет принадлежать L_2 т.к. оно не удовлетворяет лемме о накачке.

2.3 $L_3 = \{w \mid |w|_a = |w|_b\}$

Приведем доказательство нерегулярности через выполнение отрицания Леммы о Накачке. Само отрицание выглядит так:
$\forall p \ \ \exists \omega \in L : |\omega| \geq$ и $\forall$разбиения $\omega = xyz : \ \ |xy| \leq p$ и $y \neq \varepsilon$
Выполнено $\exists i \geq 0 \ xy^iz \notin L; \ \ \Rightarrow L \notin REG$
Теперь конкретное применение:
$\forall p \ \ \exists \omega = a^n b^n \in L_3 : |\omega| \geq$ и $\forall$разбиения $\omega = xyz : \ \ |xy| \leq p$ и $y \neq \varepsilon$
Выполнено $\exists i = 0 \ xz \notin L; \ \ \Rightarrow L \notin REG$
теперь пояснение, почему xz $\notin L$;. Т.к. $|xy| \leq p$ то в xy содержатся только а. И тогда пусть $x = a^k; \ \ y = a^l \ : \ l \geq 1; \ \ k + l \leq p$; То тогда $z = a^{p-l-k}b^p$; и $xz = a^k a^{p-k-l} b^p = a^{p-l}b^p$ т.е. слово xz содержит не равное число букв а и b. А по предположению о регулярности L_3 xz принадлежит L_3, состоящему из слов где число букв a и b равно

2.4 $L_4 = \{w \mid |w|_{ab} = |w|_{ba}\}$

Ответ да. Приведём ДКА, принимающий этот язык:



Докажем одну ветку, а вторая будет доказываться точно так же, только поменяв местами а и b.
$\Sigma = \{a, b\}; \quad Q = \{q_0, q_1, q_2, q_3, q_4\} \quad q_0 = q_0$; переходы видны на рисунке. Для определённости, первой буквой принятого слова будет а(пустое слово так же принимается просто по построению). Будем доказывать по индукции по длине слова, что принимаются слова только из языка L_4.

Пока число $|w|_{ab} = |w|_{ba} = 0$.
Шаг: При переходе по a мы остаёмся в q_1 и $|w|_{ab}$ как и $|w|_{ba}$ не меняется. При переходе по b мы попадаем в q_3, и тогда $|w|_{ab}$ увеличивается на 1. При переходах по b мы остаёмся в q_3, число подслов не меняется. И потом, перейдя по a в q_1 у нас на 1 увеличивается $|w|_{ba}$. В итоге снова $|w|_{ab} = |w|_{ba}$ и мы оказались в q_1 и оно принимающее. таким образом мы получили, что эта часть автомата принимает только слова начинающиеся на a, у которых $|w|_{ab} = |w|_{ba}$. Вторая часть ДКа работает точно так же, но со словами, начинающимися на b.
А по доказанному на семинаре, в листочке и в книге Дж. Хопкрофт, Р. Мотвани, Дж. Ульман, ДКА это способ задания регулярного языка. Так что мы доказали, что $L_4 = \{w \mid |w|_{ab} = |w|_{ba}\}$ это регулярный язык.