Формальные языки, занятие 2

Термины

Говорим, что язык L отделяет слова X и Y, если существует слово Z такое, что среди строк XZ и YZ одна распознаётся, а другая не распознаётся. Пример: если дан автомат, который распознаёт только строку `0101`, то X = `01` и Y = `010` отделимы: можно взять строку Z = `01`, и тогда XZ = `0101` будет распознано, а YZ = `01001` не будет.

Если строки X и Y не отделимы языком L, то можно написать `X ≡_L Y`.

Индексом L называют супремум размеров множеств, в которых все строки попарно отделимы языком L. Иными словами, если можно построить сколь угодно большое такое множество, то индекс бесконечный, а если максимальный размер такого множества равен k, то индекс равен k. Пример: для языка, распознающего только строк `0101`, индекс равен 5. Максимального размера множество, в котором все строки попарно отделимы, например, такое: {`0, 01, 010, 0101, 1`}.

Разбор задач

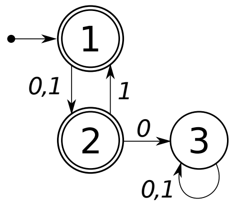
Язык слов, на чётных позициях которых стоит единица

Вводим три состояния:

1. Состояние, в котором ещё есть шанс распознать слово и следующий символ будет на нечётной позиции.
2. Состояние, в котором ещё есть шанс распознать слово и следующий символ будет на чётной позиции.
3. Состояние, где строка уже обречена.

Начальное состояние №1; терминальных состояний два: №1 и №2.

Переходы: из состояния №1 в состояние №2 можно перейти по любому символу, поскольку нам всё равно, что на нечётных позициях; из состояния №2 можно вернуться в состояние №1 по символу 1, поскольку он ожидаем на чётных позициях; по символу 0 же попадаем в состояние №3, из которого уже нельзя будет выбраться.



Язык развёрнутых слов языка из предыдущей задачи

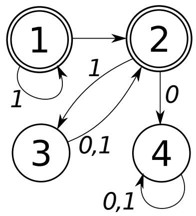
Решение сильно более сложное, хотя задача интуитивно кажется почти идентичной предыдущей. Дело в том, что в момент, когда считана строка, например, `011101`, нельзя знать, какие её позиции в итоге будут чётными с конца: либо `_1_1_1`, либо `0_1_0_`. Таким образом, нужно помнить в каждом состоянии две вещи: есть ли шанс, что данная строка будет распознана, если осталось считать ещё чётное число символов, и есть ли шанс, если осталось нечётное.

Введём четыре состояния:

1. Состояние, где нет разницы, сколько осталось символов, у строки всё равно есть шанс быть распознаной.
2. Состояние, где у строки есть шанс быть распознаной, если осталось чётное число символов.
3. Состояние, где у строки есть шанс быть распознаной, если осталось нечётное число символов.
4. Состояние, в котором уже ясно, что строка не будет распознана.

Начальное состояние №1; терминальные – №1 и №2 (так как 0 символов тоже является чётным числом).

Переходы: №1 переходит в себя по единице, поскольку префикс, состоящий из единиц, можно выкинуть из строки, не повлияв на то, будет ли она распознана. По нулю №1 переходит в №2; если такое происходит, то была прочитана строка вида `1*0`, и если осталось нечётное число символов, то уже ничего не спасти – ноль встретится на чётной с конца позиции – но если осталось чётное число символов и вся оставшаяся строка имеет вид `(1[01])*`, то она всё ещё будет распознана. В любом ином случае она не будет распознана. `1[01]` считывается проходом из №2 по единице в №3, откуда по любому символу можно снова попасть в №2; если же в №2 пришёл 0, то строка уже не будет распознана ни в каком случае, так как в любом случае на чётной позиции будет ноль, и тогда мы попадаем в состояние №4, откуда уже не выбраться.



Индекс языка слов над алфавитом {0, 1}, содержащих подстроку 01

Чтобы понять, какой индекс этого языка, разобьём все существующие строки на три класса: * Строки, в которых не встречался ноль (строки вида 1*); * Строки, в которых встречались ноли, но за ними не было единиц (`1*0+`); * Строки, в которых встречался ноль, за которым идёт единица (`1*0+1[01]*`).

Строк, не входящих ни в один из этих классов, не существует.

Эти классы примечательны тем, что, во-первых, внутри одного класса находятся только попарно неотделимые языком L слова, а во-вторых, слова разных классов всегда попарно отделимы L. Если эти два утверждения получится доказать, то задача решена: раз за пределами этих классов строк не существует, то можно найти как максимум три попарно отделимых слова (например, {1, 10, 101}) и индекс равен 3.

1. Если в строке не встречался ноль, это значит, что эти строки просто состоят из наборов единиц. Пусть у нас есть такие строки X и Y. Выберем теперь строку Z такую, что XZ распознается. Так как X состоит из набора единиц, чтобы XZ распознался, в Z должна быть подстрока 01. Но тогда YZ тоже распознается.

Возьмём теперь строки X и Y, каждая из которых имеет вид `1*0+` (второй класс). Чтобы XZ распознался, надо, чтобы в Z встречалась единица. Но тогда YZ тоже распознается.

Наконец, если строки X и Y относятся к третьему классу, то что к ним ни приписывай, результат всё равно распознается.

2. Слово X первого класса и слово Y второго класса отделимы строкой Z = 1. Слово XZ принадлежит первому классу и не распознается, а слово YZ принадлежит третьему и распознается.

Слово X третьего класса и слово Y любого другого отделимы по пустой строке (Z = ""): X распознается, а Y не распознается.

Индекс языка слов вида $a^n b^m$, где n ≡ m mod 3

Разобьём все строки на семь классов: * `(aaa)*` * `(aaa)*a` * `(aaa)*aa` * `a{n}b{m}`, где $m > 0$ и n ≡ m mod 3 * `a{n}b{m}`, где $m > 0$ и n + 1 ≡ m mod 3 * `a{n}b{m}`, где $m > 0$ и n + 2 ≡ m mod 3 * `a*b+a[ab]*` (строки, которые не имеют вид $a^n b^m$)

Представители разных классов всегда попарно отделимы, а представители одного класса неотделимы. Рассуждения не сложнее, чем в предыдущей задаче, но их больше.