Курсовая работа

Методические указания по выполнению

Курсовая работа должна выполняться после изучения всего теоретического материала и выполнения лабораторных работ. Курсовая работа состоит в написании программы в соответствии с заданием. В случае наличия ошибок в программе она возвращается на доработку.

**При выполнении работы над ошибками необходимо сохранять замечания преподавателя, а изменения в отчёт вносить другим цветом.**

Выполнение задания включает разработку программного средства, тестирование его на наборе данных и написание отчёта по работе.

Отчёт должен содержать:

1. титульный лист;
2. номер варианта и текст задания;
3. описание алгоритма решения задачи с иллюстрацией его на своём примере;
4. описание основных блоков программы;
5. текст программы;
6. результаты тестирования программы;
7. распечатку файла результатов – 2–3 примера.

Высылаемый на проверку преподавателю архив должен содержать отчёт и все файлы программы – все модули, exe-модуль, файлы данных, файлы результатов.

**Внимание!!!**

**Необходимо предусмотреть обработку ошибок.   
Никакие входные данные не должны нарушать работу программы!! Программа не должна «зависать» или прекращать выполнение по неизвестной причине – обязательна выдача соответствующей диагностики.**

При выполнении задания следует использовать средства объектно-ориентированного программирования. Выбор конкретного средства разработки оставляется за студентом. При наличии определённых требований к программному обеспечению студенту следует уведомлять об этом преподавателя.

Рекомендуется при разработке программного средства использовать материалы лабораторных работ (в зависимости от темы задания).

Программа должна управляться посредством меню, в котором должны присутствовать следующие пункты: "Автор", "Тема" (с полной информацией о разработчике и теме задания), "Данные" (выбор способа задания исходных данных – чтение из файла или ввод с клавиатуры), "Расчёты", "Запись результатов в файл" – и другие, определяемые конкретным заданием. При вводе данных с клавиатуры необходимо использовать соответствующую форму, а также предусмотреть возможность вызова справки с примером формата данных. При чтении из файла – должна открываться своя папка. Все результаты расчётов должны отображаться на экране и выводиться в файл (по требованию пользователя). При введении автором каких-либо ограничений (размер алфавита и т.п.) они должны быть описаны в пояснительной записке и в соответствующем пункте меню.

## Вариант 1

Написать программу для автоматического построения детерминированного конечного автомата (ДКА) по словесному описанию языка.

Вход программы: алфавит языка, обязательная начальная подцепочка, кратность длины всех цепочек языка, цепочки для распознавания.

Выход: построенный ДКА (все 5 элементов), результат проверки цепочек.

Подробно:

Язык задан своим алфавитом, обязательной начальной цепочкой и указанием кратности длины всех цепочек языка. В начальной цепочке не должно находиться символов, не содержащихся в алфавите. В крайнем случае она может быть и пустой. Кратность длины цепочек задаётся любым натуральным числом. Кратность, равная единице, равносильна отсутствию ограничений на длину цепочек.

Программа должна:

1. по предложенному описанию регулярного языка строить ДКА, распознающий этот язык, в том виде, как он рассматривался в теории, раздел 2.2.2;
2. с помощью построенного ДКА проверять вводимые пользователем цепочки на их принадлежность этому языку.

ДКА может быть полностью или неполностью определённым – по желанию разработчика. Функция переходов ДКА может изображаться в виде таблицы и графа, вариант вида отображения должен выбираться в меню.

После построения ДКА пользователь может вводить произвольные цепочки для проверки их на принадлежность исходному языку. Разбор цепочек автоматом следует поэтапно отображать на экране в виде последовательной смены конфигураций в соответствии с лабораторной работой №2.

Рассмотрим пример построения ДКА.

Задан язык: алфавит {a,b,c}, обязательная начальная цепочка ‘aab’, длина цепочек кратна 2. Анализируем задание: язык будет состоять из цепочек чётной длины, начинающихся с ‘aab’, например {aaba, aabb, aabc, aabaaa, aabaab, aabccc,…}. Тогда ДКА должен иметь вид M(Q,{a,b,с},δ,q0,F), множество состояний Q и заключительные состояния F определятся в процессе построения. Разберёмся с построением функции переходов. Очевидно, что пустая цепочка в языке не содержится (поскольку есть непустая обязательная начальная цепочка). Сначала определимся с минимальной цепочкой языка и построим для неё граф переходов. Это ‘aab’ и еще один (любой) символ алфавита.

q0

q1

q4

b

a

q2

q3

a

a,b,c

После этой цепочки допускается её наращивание парами любых символов алфавита. Возврат из последнего состояния в q3 по любому символу алфавита позволит добавлять такие пары символов. Функция переходов δ построена:

q0

q1

q4

b

a

q2

q3

a

a,b,c

a,b,c

Q={q0,q1,q2,q3,q4}, F={q4}.

В табличном виде она представляется очевидным образом.

## Вариант 2

Написать программу для автоматического построения регулярной грамматики (леволинейной или праволинейной) по словесному описанию языка.

Вход программы: алфавит языка, обязательные начальная и конечная подцепочки, тип грамматики (ЛЛ либо ПЛ), 2 числа – диапазон длин для генерации цепочек.

Выход: построенная грамматика (все 4 элемента), результат генерации цепочек.

Подробно:

Язык задан своим алфавитом, обязательной начальной и конечной подцепочками, которые должны присутствовать во всех цепочках языка. В заданных подцепочках не должно находиться символов, не содержащихся в алфавите. В крайнем случае они (одна или обе) могут быть и пустыми. Особый случай – это пересечение заданных цепочек по одному или нескольким символам, в том числе и полное их совпадение.

Программа должна:

1. по предложенному описанию регулярного языка строить регулярную грамматику, генерирующую этот язык, в том виде, как она рассматривалась в теории, раздел 1.3.1;
2. с помощью построенной грамматики генерировать все цепочки языка в заданном пользователем диапазоне длин.

Грамматика должна строиться и праволинейная, и леволинейная. Желаемый тип грамматики выбирается пользователем в меню. Все правила грамматики должны соответствовать выбранному типу! Отдельно следует указывать, какой нетерминальный символ является целевым. Если в грамматике используется пустое правило, то необходимо дать пояснение, каким именно символом обозначается пустая цепочка.

После построения грамматики пользователь может убедиться в её правильности путём генерации всех цепочек языка в том диапазоне длин, который он задаст. Генерацию каждой цепочки языка следует поэтапно отображать на экране в виде цепочки вывода (в соответствии с примерами раздела 1.4.1.). Генерация осуществляется в соответствии с лабораторной работой №1.

Рассмотрим пример построения регулярной грамматики.

Задан язык: алфавит {a,b,c}, обязательная начальная цепочка ‘саа’, конечная цепочка ‘aab’. Анализируем задание: язык будет состоять из цепочек любой длины, начинающихся с цепочки ‘cаа’ и заканчивающихся на ‘aab’, например {сaab, саaab, сaaааb, caaсааb, сaabaab, …}.

Для определённости выберем праволинейную грамматику (ПЛ). В правой части её правил вывода может быть только один нетерминал, и он должен располагаться справа от цепочек терминальных символов.

Все цепочки языка должны начинаться с заданной цепочки ‘саа’, после которой может находиться всё, что угодно. Значит, правило для целевого символа будет иметь вид: S→caaА, а для нетерминала А: А→aA|bA|cA. В конце должна быть ‘aab’, значит, выход из нетерминала А будет по ней: А→aA|bA|cA|aab. Осталось учесть возможные пересечения цепочек, это ‘сaab’, ‘саaab’. Их нужно выводить из целевого символа. Итак, грамматика имеет вид: G({a,b,c},{S,A},P,S), Р: S→caaА|сaab|саaab; А→aA|bA|cA|aab.

## Вариант 3

Написать программу для автоматического построения регулярного выражения (РВ) по словесному описанию языка.

Вход программы: алфавит языка, обязательная начальная подцепочка, выбранный символ алфавита, его кратность (натуральное число), 2 числа – диапазон длин для генерации цепочек.

Выход: построенное регулярное выражение, результат генерации цепочек.

Подробно:

Язык задан своим алфавитом, обязательной начальной цепочкой и указанием кратности вхождений некоторого символа во все цепочки языка. В начальной цепочке не должно находиться символов, не содержащихся в алфавите. В крайнем случае она может быть и пустой. Кратность числа вхождений выбранного символа задаётся любым натуральным числом. Кратность, равная единице, равносильна отсутствию ограничений на вхождения данного символа, т.е. он может встречаться в цепочках любое количество раз. При любом значении кратности количество вхождений символа в цепочку может быть и нулевым.

Программа должна:

1. по предложенному описанию регулярного языка строить РВ, генерирующее этот язык, в том виде, как было рассмотрено в теории, раздел 2.1.1;
2. с помощью построенного РВ генерировать все цепочки языка в заданном пользователем диапазоне длин.

При генерации цепочек языка по РВ можно использовать любые структуры и алгоритмы, в том числе и прямой перебор.

Рассмотрим пример построения регулярного выражения.

Задан язык: алфавит {0,1,a,b}, обязательная начальная цепочка ‘01a’, количество вхождений символа ‘а’ кратно 2. Анализируем задание: язык будет состоять из цепочек с чётным количеством символов ‘а’, начинающихся с ‘01a’, например {01аa, 01аba, 01аab, 01аac, 01аaaa,…}. Для начала определим минимально возможную цепочку этого языка. Это ‘01aa’ – она удовлетворяет всем условиям и является при этом самой короткой. Это основа нашего РВ.

Удлинить эту цепочку можно, поместив после неё или между цепочкой ’01a’ и ‘a’ в ней любое количество любых символов алфавита, за исключением символа ‘a’ – для него нужно подсчитывать количество, поэтому его следует рассматривать особо. Заданную цепочку ’01a’ разрывать нельзя. Для генерации любого количества символов используется итерация, а выбор одного из нескольких символов записывается в виде суммы.

Итак, получим 01a(0+1+b)\*a(0+1+b)\*. Но по построенному таким образом выражению нельзя получить много символов ‘a’. Для увеличения количества ‘a’ можно добавить после построенного выражения (aa)\* для сохранения чётности. Но тогда нельзя будет между символами ‘a’ поместить другие символы алфавита. Для того чтобы сделать это возможным, заменим добавляемую конструкцию на ((0+1+b)\*a(0+1+b)\*a(0+1+b)\*)\*, сохранив чётность ‘a’ и добавив любое количество других символов. Итоговое регулярное выражение будет иметь вид:   
01a(0+1+b)\*a(0+1+b)\*((0+1+b)\*а(0+1+b)\*a(0+1+b))\*. Можно несколько упростить построенное выражение, убрав одну из скобок (0+1+b)\* – там, где она повторяется. Окончательный ответ: 01a(0+1+b)\*a(0+1+b)\*(а(0+1+b)\*a(0+1+b))\***.**

## Вариант 4

Написать программу для автоматического построения детерминированного конечного автомата (ДКА), эквивалентного заданной регулярной грамматике.

Вход программы: терминальный и нетерминальный алфавиты грамматики, целевой символ, правила грамматики, цепочки для распознавания.

Выход: построенный ДКА (все 5 элементов), результат проверки цепочек.

Подробно:

Язык задан регулярной грамматикой, причём она может быть не автоматного вида. При написании программы разработчику разрешается выбрать один из двух типов регулярной грамматики (ЛЛ или ПЛ) и следует информировать об этом пользователя. Терминальный алфавит грамматики может включать в себя любые символы, в нетерминальном алфавите могут использоваться заглавные буквы латинского алфавита или (на усмотрение разработчика) слова. Правила задаваемой грамматики должны соответствовать выбранному типу. Для того чтобы в исходной грамматике можно было использовать пустое правило, необходимо предусмотреть поле ввода для символа, которым пользователь может обозначить пустую цепочку.

Программа должна:

1. по заданной регулярной грамматике строить эквивалентный ДКА, распознающий этот же язык, в том виде, как он рассматривался в теории, раздел 2.2.2;
2. с помощью построенного ДКА проверять вводимые пользователем цепочки на их принадлежность этому языку.

ДКА должен распознавать язык, задаваемый исходной грамматикой, т.е. являться эквивалентной конструкцией. Функция переходов ДКА может изображаться в виде таблицы или графа, вариант вида её представления выбирается разработчиком. Для удобства построения автомата рекомендуется предварительно привести заданную грамматику к автоматному виду (в соответствии с лекционным разделом 2.2.1).

При выборе такого способа построения ДКА, когда сначала по заданной грамматике строится эквивалентный НКА, а затем он приводится к детерминированному виду, промежуточный результат в виде НКА необходимо также отображать на экране по просьбе пользователя.

После построения ДКА пользователь может вводить произвольные цепочки для проверки их на принадлежность исходному языку. Разбор цепочек автоматом следует поэтапно отображать на экране в виде последовательной смены конфигураций в соответствии с лабораторной работой №2.

Рассмотрим пример построения ДКА.

Язык задан праволинейной грамматикой: G({a,b,c},{S,A},P,S),   
Р: S→aA|bA|cA; А→aS|bS|cS|aab.

Сначала следует привести грамматику к автоматному виду, добавив 2 нетерминала для разделения на символы цепочки ‘aab’. Правила примут вид:   
Р`: S→aA|bA|cA; А→aS|bS|cS|aB; B→aC; C→b. Теперь нужно по правилам грамматики построить конечный автомат, в котором состояния будут получены из нетерминалов грамматики, а переходы определяются терминальными символами. Заключительным будет являться то состояние, в которое происходит переход по одному символу (C→b), следовательно, придётся добавить ещё одно состояние D. Функцию переходов сначала будем строить графически:

Как видно из графа переходов, автомат получился недетерминированный (из состояния А есть два перехода по символу ‘a’). Далее можно применить алгоритм преобразования НКА к ДКА, рассмотренный подробно в лекциях, раздел 2.2.2.

S

A

D

a

a,b,c

B

C

a

b

a,b,c

Построим таблицу переходов нашего автомата и преобразуем его в ДКА согласно вышеупомянутому алгоритму.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | ***вход*** | | | Исходную таблицу переходов отделим от остальной части жирной линией.  Для упрощения процесса будем создавать не все возможные сочетания исходных состояний, а только те, которые реально возникают при построении. Сначала занесём в таблицу SB. Затем появляются AC, SD. Состояния исходного автомата B, C, D (выделены синим) оказались недостижимыми. Удалим их. |
| ***состояние*** | | *а* | *b* | *c* |
| S | | {A} | {A} | {A} |
| A | | {S,B} | {S} | {S} |
| B | | {C} | – | – |
| C | | – | {D} | – |
| D | | – | – | – |
| SB | B | {AC} | {A} | {A} |
| AC | C | {SB} | {SD} | {S} |
| SD | D | {A} | {A} | {A} |

Новые состояния для удобства переобозначим B, C, D. Заключительным состоянием станет состояние D (поскольку в состояние SD входит то состояние D, которое было заключительным в исходном автомате). Новая таблица переходов примет следующий вид:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***вход*** | | | Построение графа переходов по таблице:  b,c  S  A  D  a  a,b,c  B  C  a  b  b,c  a  c  a,b,c |
| ***состояние*** | *a* | *b* | *c* |
| S | {A} | {A} | {A} |
| A | {B} | {S} | {S} |
| B | {C} | {A} | {A} |
| C | {B} | {D} | {S} |
| D | {A} | {A} | {A} |

Итак, ДКА построен и имеет вид: M({S,A,B,C,D},{a,b,с},δ,S,{D}), где функция переходов δ получена выше.

## Вариант 5

Написать программу для автоматического построения регулярной грамматики (леволинейной или праволинейной), эквивалентной заданному конечному автомату (КА).

Вход программы: КА в виде: алфавит языка, множество состояний, начальное состояние, множество заключительных состояний, функция переходов в табличном виде; тип грамматики (ЛЛ либо ПЛ), 2 числа – диапазон длин для генерации цепочек.

Выход: построенная грамматика (все 4 элемента), результат генерации цепочек.

Подробно:

Язык задан конечным автоматом. Автомат может быть полностью или неполностью определённым, детерминированным или недетерминированным, заключительных состояний может быть несколько.

Программа должна:

1. по предложенному распознавателю регулярного языка в виде КА строить регулярную грамматику, генерирующую этот язык, в том виде, как она рассматривалась в теории, раздел 1.3.1;
2. с помощью построенной грамматики генерировать все цепочки языка в заданном пользователем диапазоне длин.

Грамматика должна строиться и праволинейная, и леволинейная. Желаемый тип грамматики выбирается пользователем в меню. Все правила грамматики должны соответствовать выбранному типу! Отдельно следует указывать, какой нетерминальный символ является целевым. Если в грамматике используется пустое правило, то необходимо дать пояснение, каким именно символом обозначается пустая цепочка.

После построения грамматики пользователь может убедиться в её правильности путём генерации всех цепочек языка в том диапазоне длин, который он задаст. Генерацию каждой цепочки языка следует поэтапно отображать на экране в виде цепочки вывода (в соответствии с примерами раздела 1.4.1.). Генерация осуществляется в соответствии с лабораторной работой №1.

Рассмотрим пример построения регулярной грамматики.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***вход*** | | | Задан ДКА M({S,A,B},{0,1,2},δ,S,{B}), функция переходов δ представлена в табличном виде. Построить праволинейную грамматику. В правой части её правил вывода может быть только один нетерминал, и он должен располагаться справа от терминальных символов. |
| ***состояние*** | 0 | 1 | 2 |
| S | {S} | {A} | – |
| A | {B} | – | {B} |
| B | {A} | – | {A} |

Нетерминалами будут являться состояния ДКА: {S,A,B}, заключительному состоянию должно соответствовать пустое правило. Целевым символом будет нетерминал, соответствующий начальному состоянию исходного ДКА. Правила грамматики строятся по строкам таблицы переходов следующим образом: если Х∈δ(Y,t), X,Y∈Q, t – входной символ, то соответствующее правило праволинейной грамматики будет иметь вид Y → tX.

Итак, грамматика имеет вид: G({0,1,2},{S,A,B},P,S), Р: S→0S|1А; A→0B|2B; B→0A|2A|λ.

## Вариант 6

Написать программу для автоматического построения грамматики, эквивалентной заданному регулярному выражению (РВ).

Вход программы: регулярное выражение в виде строки символов, 2 числа – диапазон длин для генерации цепочек.

Выход: построенная грамматика (все 4 элемента), результат генерации цепочек.

Подробно:

Язык задан регулярным выражением. При его записи могут быть использованы символы алфавита языка, а также: «+» (выбор одного из слагаемых), круглые скобки, «\*» для обозначения итерации.

Программа должна:

1. по предложенному регулярному выражению строить эквивалентную грамматику, генерирующую этот же язык, в том виде, как она рассматривалась в теории, раздел 1.3.1;
2. с помощью построенной грамматики генерировать все цепочки языка в заданном пользователем диапазоне длин.

Грамматика может строиться любая – контекстно-свободная или регулярная, по выбору разработчика. Отдельно следует указывать, какой нетерминальный символ является целевым. Если в грамматике используется пустое правило, то необходимо дать пояснение, каким именно символом обозначается пустая цепочка.

После построения грамматики пользователь может убедиться в её правильности путём генерации всех цепочек языка в том диапазоне длин, который он задаст. Генерацию каждой цепочки языка следует поэтапно отображать на экране в виде цепочки вывода (в соответствии с примерами раздела 1.4.1.). Генерация осуществляется в соответствии с лабораторной работой №1.

Рассмотрим пример построения КС-грамматики.

Задано регулярное выражение: ((0+1+b)\*a(0+1+b)\*a)\*(0+1+b)\*a(0+1+b)\*01a.

Для построения правил грамматики следует сделать разбор исходного регулярного выражения. Каждая скобка обозначается своим нетерминалом. Если на скобке стоит звёздочка (итерация), значит, на этом нетерминале будет явная рекурсия и пустое правило. Если в выражении стоит «+», то это означает альтернативу в правилах.

Обозначим первую большую скобку через A=((0+1+b)\*a(0+1+b)\*a)\*, вторую B=(0+1+b)\*. Само РВ должно порождаться из целевого символа грамматики. Тогда первое правило будет иметь вид: S→ABaB01a. В правиле для A будет присутствовать B: A→BaBa. Поскольку на скобке есть звёздочка, то надо добавить рекурсию и пустое правило: A→BaBaA|λ. Нетерминал B рекурсивно порождает любые символы, кроме ‘a’: B→0B|1B|bB|λ. Итак, грамматика построена, выпишем её полностью.

G({0,1,a,b},{S,A,B},P,S), где P: S→ABaB01a; A→BaBaA|λ, B→0B|1B|bB|λ.

Аналогично строится регулярная грамматика, только там следует учитывать, что в правой части правил может использоваться не более одного нетерминала, и располагаться во всех правилах грамматики он должен с одной стороны от цепочки терминальных символов.

## Вариант 7

Написать программу для автоматического приведения заданной контекстно-свободной грамматики (КС-грамматики) к нормальной форме Хомского (БНФ).

Вход программы: терминальный и нетерминальный алфавиты грамматики, целевой символ, правила грамматики, 2 числа – диапазон длин для генерации цепочек.

Выход: построенная грамматика в БНФ (все 4 элемента), результат генерации цепочек по обеим грамматикам.

Подробно:

Язык задан КС-грамматикой, причём для приведения к БНФ она должна находиться в каноническом виде (раздел лекций 3.2.2). Приводить её к этому виду не требуется, достаточно только проверить корректность задания – действительно ли исходная грамматика находится в каноническом виде – и при отрицательном результате выдать соответствующее сообщение. Причём в этом сообщении должны быть конкретно указаны причины, почему именно грамматика не имеет канонического вида (например: «в грамматике присутствуют цепные правила: А→С», или «в грамматике присутствует недостижимый символ: В»). Для того чтобы в исходной грамматике можно было использовать пустое правило, необходимо либо предусмотреть поле ввода для символа, которым обозначается пустая цепочка, либо дать пояснения пользователю, как именно ему следует задавать пустое правило.

Программа должна:

1. проверить заданную КС-грамматику – находится ли она в каноническом виде и при отрицательном результате выдать сообщение;
2. привести заданную КС-грамматику к нормальной форме Хомского (раздел 3.3.1);
3. проверить построенную грамматику (БНФ) на эквивалентность исходной.

Для проверки построенной грамматики в БНФ на эквивалентность исходной по обеим грамматикам следует сгенерировать множества всех цепочек в заданном пользователем диапазоне длин и проверить эти множества на идентичность. Генерация цепочек осуществляется в соответствии с лабораторной работой №1. При обнаружении несовпадения должна выдаваться диагностика различий – где именно несовпадения и в чём они состоят. Для удобства сравнения множества цепочек необходимо упорядочить, цепочки перенумеровать, для генерации цепочек по каждой грамматике сделать отдельную кнопку. Следует предусмотреть возможность многократного изменения диапазона длин цепочек и повторной их генерации для новых значений длины.

Приведение КС-грамматики к нормальной форме Хомского следует осуществлять согласно алгоритму, изложенному в лекционном материале (раздел 3.3.1).

## Вариант 8

Написать программу для автоматического построения детерминированного конечного автомата (ДКА) по словесному описанию языка.

Вход программы: алфавит языка, обязательная конечная подцепочка, цепочки для распознавания.

Выход: построенный ДКА (все 5 элементов), результат проверки цепочек.

Подробно:

Язык задан своим алфавитом и обязательной конечной подцепочкой всех цепочек языка. В конечной подцепочке не должно находиться символов, не содержащихся в алфавите. В крайнем случае она может быть и пустой.

Программа должна:

1. по предложенному описанию регулярного языка строить ДКА, распознающий этот язык, в том виде, как он рассматривался в теории, раздел 2.2.2;
2. с помощью построенного ДКА проверять вводимые пользователем цепочки на их принадлежность этому языку.

ДКА должен быть полностью определённым. Функция переходов ДКА может изображаться в виде таблицы или графа, вариант вида её представления выбирается разработчиком.

Наиболее простой способ построения такого ДКА состоит в том, чтобы сначала по описанию языка построить НКА (недетерминированный конечный автомат), а затем преобразовать его согласно рассмотренному в разделе 2.2.2 алгоритму. При выборе такого способа построения ДКА промежуточный результат в виде НКА необходимо также отображать на экране по просьбе пользователя.

По желанию автора допускаются и другие способы построения ДКА.

После построения ДКА пользователь может вводить произвольные цепочки для проверки их на принадлежность исходному языку. Разбор цепочек автоматом следует поэтапно отображать на экране в виде последовательной смены конфигураций в соответствии с лабораторной работой №2.

Рассмотрим пример построения ДКА.

Задан язык: алфавит {a,b,c}, обязательная конечная подцепочка ‘aaba’. Анализируем задание: язык будет состоять из цепочек любой длины, заканчивающихся на ‘aaba’, например {aaba, аaaba, baaba, caaba, aaaaba, …}. Тогда ДКА должен иметь вид M(Q,{a,b,с},δ,q0,F), множество состояний Q и заключительные состояния F определятся в процессе построения. Разберёмся с построением функции переходов δ. Очевидно, что пустая цепочка в языке не содержится (поскольку есть непустая обязательная конечная цепочка). Сначала определимся с минимальной цепочкой языка – это ‘aaba’, и построим для неё граф переходов.

q0

q1

q4

b

a

q2

q3

а

a

Если выбрать способ с предварительным построением НКА, то такой автомат выглядит очевидным образом. Сначала могут быть прочитаны любые символы алфавита в любом количестве, а затем конечная подцепочка:

q0

q1

q4

b

a

а,b,с

q2

q3

а

a

Недетерминированность автомата вызвана тем, что из начального состояния существует два перехода по одному символу алфавита (‘a’). Осталось преобразовать построенный автомат в детерминированный. Для этого построим таблицу переходов:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | ***вход*** | | | Исходную таблицу переходов отделим от остальной части жирной линией.  Для упрощения процесса будем создавать не все возможные новые состояния, которые могут получиться в результате сочетаний исходных состояний, а только те, которые реально возникают при построении. Сначала это единственное состояние q0q1 – занесём его в таблицу. Затем последовательно появятся q0q1q2, q0q3, q0q1q4. Все состояния исходного автомата, кроме q0, оказались недостижимыми. В таблице они выделены синим цветом. Удалим их. |
| ***состояние*** | | a | b | c |
| q0 | | {q0,q1} | {q0} | {q0} |
| q1 | | {q2} | – | – |
| q2 | | – | {q3} | – |
| q3 | | {q4} | – | – |
| q4 | | – | – | – |
| q0q1 | A | {q0q1q2} | {q0} | {q0} |
| q0q1q2 | B | {q0q1q2} | {q0q3} | {q0} |
| q0q3 | C | {q0q1q4} | {q0} | {q0} |
| q0q1q4 | D | {q0q1q2} | {q0} | {q0} |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***вход*** | | | Новые состояния для удобства переобозначим A, B, C, D. Заключительными состояниями станут те, которые содержат q4. Здесь такое состояние одно – D. Новая таблица переходов представлена слева:  c  a  q0  A  D  b  a  b,с  B  C  a  a  b,c  a  b,c  b,c |
| ***состояние*** | a | b | c |
| q0 | {A} | {q0} | {q0} |
| A | {B} | {q0} | {q0} |
| B | {B} | {C} | {q0} |
| C | {D} | {q0} | {q0} |
| D | {B} | {q0} | {q0} |

Граф переходов построен по таблице:

Q={q0,A,B,С,D }, F={D}.

ДКА построен.

## Вариант 9

Написать программу для автоматического построения регулярного выражения (РВ) по словесному описанию языка.

Вход программы: алфавит языка, обязательные начальная и конечная подцепочки, кратность длины всех цепочек языка, 2 числа – диапазон длин для генерации цепочек.

Выход: построенное регулярное выражение, результат генерации цепочек.

Подробно:

Язык задан своим алфавитом, обязательной начальной и конечной подцепочками и указанием кратности длины всех цепочек языка. В заданных подцепочках не должно находиться символов, не содержащихся в алфавите. В крайнем случае они (одна или обе) могут быть и пустыми. Особый случай – это пересечение заданных цепочек по одному или нескольким символам, в том числе и полное их совпадение. Кратность длины цепочек задается любым натуральным числом. Кратность, равная единице, равносильна отсутствию ограничений на длину цепочек.

Программа должна:

1. по предложенному описанию регулярного языка строить РВ, задающее этот язык, в том виде, как было рассмотрено в теории, раздел 2.1.1;
2. с помощью построенного РВ генерировать все цепочки языка в заданном пользователем диапазоне длин.

При генерации цепочек языка по РВ можно использовать любые структуры и алгоритмы, в том числе и прямой перебор.

Рассмотрим пример построения регулярного выражения.

Задан язык: алфавит {a,b,c}, обязательная начальная цепочка ‘саа’, конечная цепочка ‘aab’, длина цепочек кратна 2. Анализируем задание: язык будет состоять из цепочек чётной длины, начинающихся с цепочки ‘cаа’ и заканчивающихся на ‘aab’, например {сaab, сaaааb, caaасааb, сaabсaab, …}.

Для начала определим минимально возможную цепочку этого языка. Это ‘сaab’ – она получается при максимальном пересечении начальной и конечной цепочек, удовлетворяет всем условиям и является при этом самой короткой. Удлинить её можно путём уменьшения области пересечения начальной и конечной цепочек. Следующая по длине цепочка могла бы быть ‘сaаab’ (пересечение по одному символу), но она не удовлетворяет условию чётности длины. А вот ещё следующая цепочка – ‘сaaааb’ – получена простым соединением начальной и конечной цепочек. Это основа нашего РВ. Её можно удлинять, помещая между начальной и конечной цепочками некоторое количество символов алфавита, причём для сохранения чётности длины это количество тоже должно быть чётным. Для генерации любого количества пар символов используется итерация, а выбор одного из нескольких символов записывается в виде суммы.

Итоговое регулярное выражение будет иметь вид: саа((а+b+c)(a+b+c))\*aab+caaaab.

Если в этом же задании изменить кратность длины на единицу, то итоговое РВ изменится таким образом: добавлять можно будет не пары символов, а по одному символу любое количество раз, и пересечения цепочек будут использованы все. Итоговое регулярное выражение будет иметь вид: саа(a+b+c)\*aab+caaab +caaaab.

## Вариант 10

Написать программу для автоматического построения регулярной грамматики (леволинейной или праволинейной) по словесному описанию языка.

Вход программы: алфавит языка, обязательная конечная подцепочка, кратность длины всех цепочек языка, тип грамматики (ЛЛ либо ПЛ), 2 числа – диапазон длин для генерации цепочек.

Выход: построенная грамматика (все 4 элемента), результат генерации цепочек.

Подробно:

Язык задан своим алфавитом, обязательной конечной подцепочкой, которая должна присутствовать во всех цепочках языка, и указанием кратности длины всех цепочек языка. В конечной цепочке не должно находиться символов, не содержащихся в алфавите. В крайнем случае она может быть и пустой. Кратность длины цепочек задаётся любым натуральным числом. Кратность, равная единице, равносильна отсутствию ограничений на длину цепочек.

Программа должна:

1. по предложенному описанию регулярного языка строить регулярную грамматику, генерирующую этот язык, в том виде, как она рассматривалась в теории, раздел 1.3.1;
2. с помощью построенной грамматики генерировать все цепочки языка в заданном пользователем диапазоне длин.

Грамматика должна строиться и праволинейная, и леволинейная. Желаемый тип грамматики выбирается пользователем в меню. Все правила грамматики должны соответствовать выбранному типу! Отдельно следует указывать, какой нетерминальный символ является целевым. Если в грамматике используется пустое правило, то необходимо дать пояснение, каким именно символом обозначается пустая цепочка.

После построения грамматики пользователь может убедиться в её правильности путём генерации всех цепочек языка в том диапазоне длин, который он задаст. Генерацию каждой цепочки языка следует поэтапно отображать на экране в виде цепочки вывода (в соответствии с примерами раздела 1.4.1.). Генерация осуществляется в соответствии с лабораторной работой №1.

Рассмотрим пример построения регулярной грамматики.

Задан язык: алфавит {a,b,c}, обязательная конечная цепочка ‘aab’, кратность длины 2. Анализируем задание: язык будет состоять из цепочек чётной длины, заканчивающихся на ‘aab’, например {сaab, аaab, bааb, caaааb, сabaab, …}.

Для примера выберем праволинейную грамматику (ПЛ). В правой части её правил вывода может быть только один нетерминал, и он должен располагаться справа от цепочек терминальных символов. Для соблюдения кратности длины цепочек будет использоваться чередование нетерминалов. Длина кратна двум, следовательно, потребуется два нетерминала: S→aA|bA|cA, А→aS|bS|cS. Выход должен происходить по обязательной конечной цепочке ‘aab’, в силу чётности длины цепочек перед ней должен быть как минимум один символ. Следовательно, эта цепочка должна порождаться нетерминалом А: А→aab.

Итак, построенная грамматика имеет вид:   
G({a,b,c},{S,A},P,S), Р: S→aA|bA|cA; А→aS|bS|cS|aab.