

Семинар 2. Обработка символьной информации

Цель семинара: Овладение навыками обработки символьных данных.

1. Задание

1. Проработайте (еще раз) примеры 1 и 2 §2.1.3 лекции 2 «Обработка символьной информации и документов сложной структуры». Попробуйте решить эти задачи самостоятельно, не заглядывая в лекцию. Если Ваше решение отличается от лекционного, проанализируйте отличия. Как они повлияют на работу программы?
2. Отладьте программы примеров 1 и 2 лекции 2.
3. Выполните свой вариант задания (с Вашим номером по списку группы) из раздела 2. Отчет по выполнению задания делайте по образцу примера 1 §2.1.3 лекции 2. Обратите внимание, что отчет должен содержать блок-схему-алгоритма.
4. Выполните свой вариант задания (с Вашим номером по списку группы) из раздела 3. Обратите внимание, что это задание надо выполнять, самостоятельно разработав подпрограмму (функцию СИ) обработки строк и вызвав ее для каждой строки, указанной в условии, – см. пример 2 лекции 2.

Жду отчетов по индивидуальному заданию!

2. Обработка последовательности символов

Дан текст – однострочная последовательность символов. В некоторых задачах предполагается, что текст разбит на слова. Под словом понимается группа символов, не содержащая пробелов и отделенная пробелами от других слов. При решении этих задач не рекомендуется использовать библиотечные функции работы со строками, за исключением, быть может, функции определения длины строки.

1. Определить, содержит ли текст символы, отличные от букв и пробелов.
2. Все символы, предшествующие первому вхождению символа *, заменить на *.
3. Все символы, стоящие после последней точки, заменить пробелами.
4. Выяснить, является ли текст десятичной записью целого числа.
5. Удалить из текста все символы, стоящие после первого вхождения символа *a*.

6. Изменить текст так, чтобы вместо строчных букв стояли соответствующие прописные буквы, а вместо прописных — строчные. Определить общее число произведенных замен.
7. Если текст содержит только малые латинские буквы, определить, упорядочен ли он по алфавиту.
8. Определить, есть ли в тексте слова, начинающиеся с заданного символа.
9. Определить, содержит ли букву *A* первое слово заданного текста.
10. Проверить, имеет ли место соответствие открывающихся и закрывающихся круглых скобок.
11. Выяснить, является ли текст идентификатором. Пояснение: идентификатор — это последовательность букв, цифр и символов подчеркивания, на первом месте может стоять только буква и символ подчеркивания.
12. Удалить из текста все цифры.
13. Определить число символов, отличных от букв и пробелов, встретившихся в тексте до первой точки.
14. Вывести на экран первое слово текста, начинающееся с буквы *A*.
15. Определить, сколько раз входит в последнее слово текста заданный символ.
16. Определить, сколько слов оканчивается на заданный символ.
17. Вывести на экран первое слово текста, оканчивающееся на заданный символ.
18. Вывести на экран первое слово текста, начинающееся с сочетания *ph*.
19. Определить число символов первого слова.
20. Определить число символов последнего слова.
21. Вычислить, сколько слов встречается в тексте до первой точки.
22. Каждое слово *one* заменить на *two*.
23. Первое вхождение слова *one* заменить на *two*.
24. Последнее вхождение слова *one* заменить на *two*.
25. Определить, сколько раз в тексте встречается сочетание символов *one*.
26. Определить, есть ли в тексте слова, начинающиеся с сочетания *ph*.
27. Определить, есть ли в тексте слова, заканчивающиеся на сочетание *ed*.
28. Если в тексте встречаются подряд две точки, то последнюю точку пары заменить пробелом.
29. Удалить из текста все пробелы, стоящие в конце слова (до первого символа с конца, не являющегося пробелом).
30. Удалить из текста все пробелы, стоящие в начале слова (до первого символа, не являющегося пробелом).

3. Индивидуальное задание на разработку подпрограмм, обрабатывающих строки

Разработать программу, выполняющую однотипные действия с одной или несколькими однострочными последовательностями символов (они обозначены A , B , C), самостоятельно выделив одну или несколько подпрограмм.

Требования к подпрограммам:

- каждая подпрограмма является функционально завершённой и должна вызываться более одного раза с разными фактическими параметрами;
 - подпрограмма обработки не должна содержать ввод (вывод) данных;
 - в подпрограммах не рекомендуется использовать глобальные переменные.
1. Определить, сколько раз в последовательности A встречается буква N и сколько раз в последовательности B встречается цифра 9.
 2. Заменить в последовательности A все символы * на пробелы, а в последовательности B — все точки на запятые.
 3. Определить, содержит ли последовательность A строчные латинские буквы, а последовательность B — прописные.
 4. Определить, сколько раз в последовательности A символ : встречается до первой точки, и сколько раз в последовательности B символ ; встречается до первой запятой.
 5. Для заданной последовательности A проверить, встречается ли среди ее первых 10 символов буква s , а для последовательности B проверить, встречается ли среди ее первых 20 символов буква q .
 6. Определить номер первого вхождения заданного символа в каждую из последовательностей A , B , C .
 7. Определить, сколько слов в последовательности A оканчивается на букву N и сколько слов в последовательности B оканчивается на букву Y .
 8. Проверить, является ли каждая из последовательностей A , B , C десятичной записью целого числа.
 9. Проверить, является ли последовательность A правильным скобочным выражением. Учесть расстановку круглых, квадратных и фигурных скобок.
 10. В последовательности A все символы, стоящие после последней точки, заменить на пробелы, а в последовательности B все символы, стоящие после последней запятой, заменить на символ *.
 11. Исключить из каждой последовательности A , B , C все символы, отличные от малых латинских букв.
 12. Определить, сколько раз в последовательности A входит слово one, а в последовательности B — слово two.
 13. Проверить, входит ли в последовательности A слово one, а в последовательности B — слово two.

14. Определить, сколько слов в последовательности A начинается с буквы N и сколько слов в последовательности B начинается с буквы Y .
15. Проверить, есть ли в последовательности A слова, начинающиеся с буквы N , и есть ли в последовательности B слова, начинающиеся с буквы Y .
16. Определить число слов, встречающихся в последовательности A до первой точки, и число слов, встречающихся в последовательности B до первой запятой.
17. Для каждой из последовательностей A, B, C выяснить, имеется ли в них пара соседних одинаковых символов.
18. Из последовательности A удалить все символы, стоящие после последней точки, а из последовательности B удалить все символы, стоящие после последнего восклицательного знака.
19. Для каждой из последовательностей A, B, C вычислить, сколько символов расположено от последней точки до конца этих последовательностей.
20. Определить, в какой из двух последовательностей, A или B , последнее слово имеет наибольшую длину.
21. Для каждой последовательности, A и B , выяснить, какое из слов – первое или последнее – имеет большую длину.
22. В каждой из последовательностей A, B, C заменить круглые скобки на соответствующие (открывающие или закрывающие) квадратные.
23. В каждой из последовательностей A, B, C заменить малые латинские буквы одноименными большими.
24. Выделить первое слово в каждой из последовательностей A, B, C .
25. Для последовательности A определить частоту вхождения каждой из русских гласных букв ($a, e, и, o, y, э, ю, я$), а для последовательности B определить частоту вхождения символов запятая и точка.
26. В каждой из последовательностей A, B, C переставить символы таким образом, чтобы первый символ стал последним, второй предпоследним, и т. д.
27. Выяснить, является ли каждая из последовательностей A, B, C идентификатором. Пояснение: идентификатор — это последовательность букв, цифр и символов подчеркивания, на первом месте может стоять только буква и символ подчеркивания
28. Вычислить число слов в каждой из последовательностей A, B, C .
29. Из последовательностей A, B, C удалить все пробелы.
30. Проверить, входит ли в последовательность A группа символов *one*, а в последовательность B — группа символов *two*.