

Лабораторная работа N2

Цель научить использовать класс RegEx для анализа текстовых файлов и разрабатывать лексические анализаторы.

Требования к оформлению:

титульный лист

текст с описанием программы (WORD 2010), который должен включать:

- с описанием программы
- исходный текст программы
- дать объяснение входящих символов в регулярное выражение
представление запрограммированного регулярного выражения в виде
теоретического регулярного выражения
- Описание контрольного примера
- Результаты тестирования

rar-архив с проектом и контрольными примером.

Контрольный пример должен находится в директории test

имя WORD 2010 файла и архива латиницей по шаблону: группа-номер-фамилия.

выслать на почту 2006mag@mail.ru

Задание

Подготовить текст (10< страницы), в котором должны быть фрагменты текста, удовлетворяющие шаблону А (см таблицу).

Написать программу, которая:

1. определяет наличие фрагментов в тексте шаблонами А;
2. выдавала бы XML файлы в котором причислены фрагменты типа А

N	Шаблон А	XML файл
1.	Выделить из текста московские стационарные телефоны.	<ph></ph>
2.	Выделить из текста слова.	<w > слово</w>
3.	Выделить из текста слова с двух сторон отделенные знаками препинания.	<w > слово</w>
4.	Выделить из текста предложения с прямой речью.	<Snt > прямая речь</Snt>
5.	Выделить из текста Время форматов чч:мм:сс и чч:мм	<time></ time >
6.	Выделить из текста адреса сайтов	<a>
7.	Выделить из текста адреса электронной почты	<mail> адреса электронной почты </mail >
8.	Выделить из текста почтовые адреса	<post> адреса </post >
9.	Выделить из текста слова с корнем Маш** подготовить	<name> Маш**</name>
10	Выделить из текста слова с корнем ЮлХХ подготовить	<inx></inx>
11	Выделить из текста Почтовые индексы <txt></txt> <num></ num>	
12	Выделить из текста числа дробные	txt></txt> <inx></inx>
13	Выделить из текста МАК адреса	<mac></mac>
14	Выделить из текста Москв** Нижний Новгород*	<mac></mac>
15	Выделить из текста IP адрес <txt></txt>	<ip></ip>
16	Выделить из текста Имена файлов и отдельно имена файла XXX.docx	<file></file> <doc></doc>

17	Выделить из текста ИНН	<inn></inn>
18	Выделить из текста УДК 999.99	<udk></udk>
19	Выделить из текста Номера паспортов	<pas></pas>
20	Выделить из текста Номерf машин и отдельно машины московского региона. подготовить XML	<nm></nm> -номер машины <umm></umm> -номер машины МО
21	Из текст удалить все союзы <ntext></ntext> -номер машины	<txt>текст без союзов</txt>
22	Из текст удалить все знаки препинания	<ntext>текст без знаков препинания</ntext>
23	Из текст удалить все слов “чемодан” во всех склавнениях	
24	В тексте заменить все слова “чемодан” во всех сравнениях, словом коробка с учетом склонения	<txt></txt> <ntext></ntext> -номер машины

2.Построить Lex для языков, выполняющий следующие функции:

- идентификация лексем
- создание таблицу идентификаторов
- отбрасывание комментариев
- нормализовать пробелы
- поиск ошибок с указанием их местом

Программа должна быть многопоточной

1. C,
2. C++,
3. C#,
4. JAVA,
5. TSQL,
6. PYTHON,
7. Linux shell ,
8. windows shell,
9. SCPI,
10. Cmake,
11. JAVA script
12. VBasic
13. VBasic script
14. Fortran
15. #F
16. Xlog

Правила оформления программы:

UML – диаграмма классов. Диаграмма объектов, .

Описать

1. Формат таблиц лексем
2. Формат выходной таблицы идентификаторов
3. Формат выходной программы
4. Описать ошибки контролируемые в лексическом анализаторе.