Цель написания программы «Подсчет слов» – лингвистический анализ текста.

Функционал программы Подсчет слов Final Version:

1. В переменную у (строчка 9 кода). Передаётся текстовый файл в формате utf-8
2. Выбираем какую аналитику (подсчет) вести – варианты с русским, латинскими буквами, диакритическими знаками и умлаутом (франц и немец языки соответственно)
3. Программа производит очитку текста от ненужных –мусорных знаков (excludes\_words = r"[#…\[(%\]!:@)–.;—,\* «»1234567890?„-]")
4. После выбора программа производит подсчет:

А) всех букв в тексте, - количество

Б) всех слов, - количество

В) повторяемость всех букв

Г) Повторяемость всех слов

Д) проверка на уникальность (Словарный запас текста) - , - количество слов

Е) располагает буквы и слова по частотности и затем по алфавиту

Ж) полученные данные формирует в базу данных в виде txt файлов: файл с

заключительной аналитикой (название файл сохраняется как «имя файла с анализируемым текст-дата-время ) Таким образом всегда видно от какого текста этот файл с аналитикой и за какое время. В аналитику выведены все полученные цифры и списки букв и слов

д) в txt файле \_time\_statistic идет сплошной статистический учет всех запросов программы и времени на исполнение этих запросов – все данные запросах сохраняются друг за другом.

Не решенные задачи:

1. определение части речи выделенного слова (существительное, глагол и проч)
2. Выявление однокоренных слов и различных форм слова (Онегин или Онегину или Онегина или Онегене = Онегин)