Задание 2. Поисковая система

Задание.
Коллекция файлов формата TXT хранит множество текстовых документов. Пользователь вводит программе запрос в виде строки печатных символов, разделенных пробелами, программа выполняет поиск по запросу пользователя в коллекции документов и выдает результат в виде списка названий (имен файлов) документов и величины их релевантности запросу пользователя.

Указания к работе.
Программа поиска должна использовать следующий общий алгоритм:
1. Выполнить лексический анализ (поиск и выделение лексем/токенов в тексте).
2. Удалить стоп-слова (списки стоп-слов для разных языков есть в сети Интернет).
3. Выполнить стемминг (выделение основы слов) токенов (использовать существующие алгоритмы и их реализации, например, стемминг Портера).
4. Использовать одну из моделей поиска по материалам лекций.