Задание 3

Одним из основных направлений деятельности «Предприятия» является генная инженерия. Все исследования в этой области проводятся в интересах государства и населения и не нарушают никаких этических норм и принципов. Так, в отделе главного генного инженера специально для людей, имеющих непереносимость к глютену был выведен сорт пшеницы, не содержащий глютен (не смотря на отсутствие глютена, мука, получающаяся из такой пшеницы, при смешивании с водой даёт эластичное и упругое тесто).

Совсем недавно сотрудники отдела главного генного инженера определили генетическую последовательность (расшифровали ДНК) дыни. Учёные заметили, что в ДНК дыни встречаются достаточно интересные последовательности нуклеотидных остатков (аденина, цитозина, гуанина, тимина), наличие которых говорит о больших перспективах использования дыни в генной инженерии.

Для начала работ по использованию дыни в генной инженерии необходимо определить, есть ли в геноме дыни последовательности нуклеотидных остатков, интересующие учёных, и если есть, то где они находятся.

Расшифрованный геном располагается в файле 2021\_SAF03\_genome. Последовательность нуклеотидных остатков представлена в виде последовательности символов A, C, T, G без пробелов.

Работники отдела главного генного инженера могли бы и сами отыскивать нужные последовательности в геноме дыни, если бы геном не был бы представлен строкой, размером 355250176 символов (файл размером 355,3 мегабайт).

Руководство «Корпорации знаний» поставило задачу написать программу, позволяющую определять, входит ли заданная последовательности нуклеотидных остатков в геном дыни, и если входит, то начиная с каких позиций.

Отдел постановки задач департамента разработки программного обеспечения сформулировал для вас как для исполнителя следующие требования в техническом задании:

1) Программа должна иметь имя 2021\_SAF03\_Фамилия, где фамилия – фамилия автора программы на латинице.

2) Программа должна отыскивать образцы в файле при помощи алгоритма Кнута-Морриса-Пратта.

3) Так как размер файла большой (355,3 мегабайт) для работы с файлом необходимо использовать проецируемые в память файлы. При таком способе работы файл не загружается полностью в оперативную память, а проецируется на адресное пространство - файлу на жестком диске ставится в соответствие определённый участок памяти (диапазон адресов оперативной памяти); при этом чтение данных из этих адресов оперативной памяти фактически приводит к чтению данных из отображенного файла (из файла на жестком диске), а запись данных по этим адресам приводит к записи этих данных в файл (на жесткий диск).

Схема работы с файлами, проецируемыми в память, следующая (упоминаемые функции содержатся в библиотеке Windows.h):

3.1) при помощи функции CreateFile создать или открыть объект ядра «файл», идентифицирующий файл на диске;

3.2) при помощи функции CreateFileMapping cоздать объект ядра «проекция файла», чтобы сообщить системе размер файла и способ доступа к нему;

3.3) при помощи функции MapViewOfFile указать системе, как спроецировать в адресное пространство процесса объект «проекция файла».

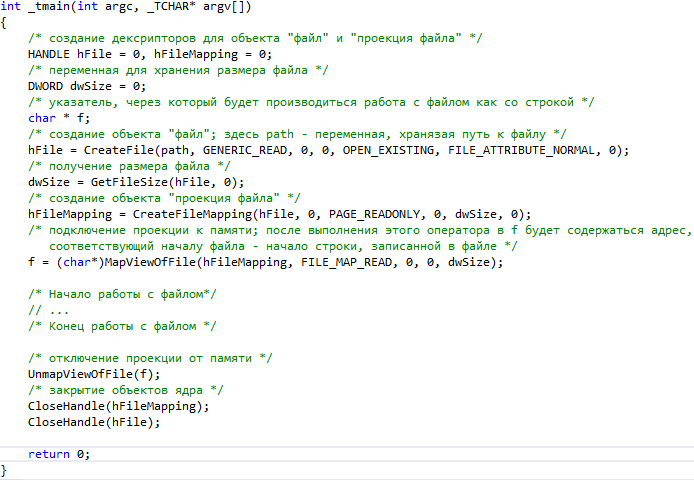
По окончании работы с проецируемым в память файлом нужно:

3.4) при помощи функции UnmapViewOfFile сообщить системе об отмене проецирования на адресное пространство процесса объекта «проекция файла»;

3.5) при помощи функции CloseHandle закрыть объект ядра «проекция файла»;

3.6) при помощи функции CloseHandle закрыть объект ядра «файл».

Фрагмент кода, который можно использовать в разрабатываемой программе:



Для получения подробной информации об используемых объектах и вызываемых функциях следует использовать библиотеку MSDN, либо книгу Пономарев В.В. Системное программирование. Учебно-методическое пособие. Редакция 05.05.2010. Озерск: ОТИ НИЯУ МИФИ, 2010. — 60 с., ил., либо дождаться дисциплины «Системное программирование».

4) Должна быть предусмотрена возможность запуска программы как с указанием параметров командной строки, так и без параметров.

5) Программа с параметрами командной строки вызывается так: 2021\_SAF03\_Фамилия имя\_файла строка

Параметр имя\_файла – указывает путь к файлу, содержащему информацию о геноме дыни. В случае если файл не найден, программа должна выводить соответствующее уведомление и завершать работу.

Параметр строка – задаёт образец, информацию о вхождении которого в файл необходимо получить.

6) В результате работы программы, вызванной с указанием параметров командной строки, создается файл 2021\_SAF03\_result, в котором в первой строке через пробел записываются допустимые сдвиги, во второй строке - количество выполненных сравнений и время, затраченное на поиск. Если нет допустимых сдвигов, в первой строке файла записывается -1, во второй строке - количество выполненных сравнений и время, затраченное на поиск.

К примеру, команда

2021\_SAF03\_Фамилия 2021\_SAF03\_genome ATTTGGGAAGGTAACCCACCTA

заставляет программу отыскать в файле 2021\_SAF03\_genome допустимые сдвиги для образца ATTTGGGAAGGTAACCCACCTA.

7) В случае, если программа вызвана без параметров, работа осуществляется в интерактивном режиме посредством меню (программа должна иметь дружественный интерфейс):

1. Загрузка файла с геномом

2. Поиск образца

3. Выход

Результат работы программы выводится в консоль.

Оценка выполнения задания производится следующим образом:

Оценка 5 (отлично): все пункты задания выполнены полностью, программа является работоспособной, исполнитель обоснованно отвечает на контрольные вопросы.

Оценка 4 (хорошо): все пункты задания выполнены полностью, программа является работоспособной, исполнитель хорошо отвечает на контрольные вопросы, но не всегда может обосновать их.

Оценка 3 (удовлетворительно): пункты задания выполнены частично, программа является частично работоспособной, исполнитель затрудняется в ответах на контрольные вопросы.

Оценка 2 (неудовлетворительно): пункты задания выполнены частично, программа не является работоспособной, исполнитель затрудняется в ответах на контрольные вопросы.

Соответствие оценок баллам балльно-рейтинговой системы при оценивании выполнения практической работы:

Оценка 5 (отлично) – 2 балла;

Оценка 4 (хорошо) – 1,5 балла;

Оценка 3 (удовлетворительно) – 1 балл;

Оценка 2 (неудовлетворительно) – 0 баллов.