

Обратите, пожалуйста, внимание: поскольку задание мы даем для того, чтобы уточнить уровень Вашей квалификации, задача формулируется не строго, оставляя как пространство для проявления сложных навыков и изобретательности, так и возможность реализовать работающее решение с помощью простых средств. Нужно реализовать настройку параметров ("обучение") кода Хаффмана на некоей базе текстов, оформленных в виде большого количества отдельных файлов. Скажем, всех файлов, лежащих в одном каталоге, полный путь к которому принимается в качестве входных данных. Файлы воспринимаются как тексты над алфавитом ASCII-символов. Процесс сбора статистики рекомендуется (не обязательно, но рекомендуется) реализовать в многопоточном формате: запускать в отдельных тредах некие "обработчики", каждый из которых будет в каждый момент времени работать с одним файлом, и по завершении своей работы каким-то образом передавать свои результаты в некую единую сущность, общую для всех обработчиков. По завершении обработки всех файлов надо выработать, собственно, код - можно в итоге представить его в виде какой-то таблицы, которую можно вывести в читаемом виде в файл или на экран.

Что такое код Хаффмана - придется разобраться по любым доступным источникам. Возможно проще это будет сделать, грубо говоря, не по википедии, а по хабрахабру. Вероятно с этого стоит начать, и потом перечитать описание задачи снова. Желательно покрыть тестами существенную долю важных частей проекта. Во всем прочем - никаких явных пожеланий нет. То есть понятно, что если, скажем, очередь обработки файлов будет реализована не полностью самостоятельно, а с помощью какого-нибудь "настоящего" брокера очередей, это будет плюсом, но не то чтобы это ожидалось. Принимаются совершенно любые варианты решения, в том числе и неполные, если Вы посчитаете, что они достаточно информативны относительно Вашей квалификации и способностей.