

Турнир по информатике

July 28, 2024

На данном турнире вам предлагается решить задачу по сжатию текста с помощью кода Хаффмана.

Чтобы победить, вы должны реализовать самое эффективное сжатие — разница между файлами до и после должна быть как можно больше при сохранении возможности получить исходный файл.

Построение кода Хаффмана сводится к построению соответствующего бинарного дерева по следующему алгоритму:

1. Составим список кодируемых символов, при этом будем рассматривать один символ как дерево, состоящее из одного элемента с весом, равным частоте появления символа в строке.
2. Из списка выберем два узла с наименьшим весом.
3. Сформируем новый узел с весом, равным сумме весов выбранных узлов, и присоединим к нему два выбранных узла в качестве детей.
4. Добавим к списку только что сформированный узел вместо двух объединенных узлов.
5. Если в списке больше одного узла, то повторим пункты со второго по пятый.

Реализация может быть на любом языке программирования, в результате должно получиться две программы: архиватор и деархиватор. В обоих случаях должен быть реализован запуск с использованием аргументов командной строки: `./compress filename`, заархивированный файл должен называться "compressed", разархивированный — "decompressed". Итоговые названия программ на ваше усмотрение.

Вам надо сдать исполняемые файлы и сам код программ (в случае python'а это одно и то же).

Часть текстов, на которых будут тестироваться ваши программы будут открытыми, мы выдадим их вам в начале.