



Домашнее задание №2 Хаффман

2 ноября 2015

Постановка задачи. Напишите программу, выполняющую сжатие двухпроходным алгоритмом Хаффмана.

Параметры командной строки.

-c	архивирование
-u	разархивирование
-f, --file <путь>	название входного файла
-o, --output <путь>	название результирующего файла

Значение параметра (если есть) указывается через пробел. Программа должна выводить на экран статистику сжатия/распаковки: размер исходных данных, размер полученных данных и размер, который был использован для хранения вспомогательных данных (например, таблицы). Не должно выводиться никакого дополнительного текста, только размеры. Все размеры в байтах.

Например:

```
$ ./huffman -c -f myfile.txt -o result.bin
15678
6172
482
```

Размер исходного файла (исходные данные): 15678 байт, размер сжатых данных (без дополнительной информации): 6172 байта, размер дополнительных данных: 482 байта. Размер всего сжатого файла: $6172 + 482 = 6654$ байта.

```
$ ./huffman -u -f result.bin -o myfile_new.txt
6172
15678
482
```

Размер распакованного файла (полученные данные): 15678 байт, размер сжатых данных (без дополнительной информации): 6172 байта, размер дополнительных данных: 482 байта. Размер всего исходного сжатого файла: $6172 + 482 = 6654$ байта.

Ограничения. Наибольший размер входного файла — 5МВ. Ограничение на время выполнения — 5 секунд.

Примечания.

- В решении не допускается использование умных указателей.
- Не выводите на экран ничего, кроме требуемой статистики, а статистику выводите ровно в том формате, который указан в условиях.

- В решении не допускаются утечки памяти.
- Открывайте файлы в двоичном режиме (`std::ios_base::binary`). В противном случае результат архивирования может оказаться не переносимым между платформами, где перевод строки представляется различным образом.
- Сжатие подразумевает представление данных в компактном виде: не следует сериализовать таблицу частот в результирующий файл в текстовом виде, или записывать сжатый алгоритмом Хаффмана поток в текстовом виде. Однако не требуется пытаться достичь максимальной эффективности сжатия, например, пытаться сжать таблицу частот при её записи в архив.
- Не храните в архиве данные, которые можно вычислить.
- Убедитесь, что ваше решение переносимо между платформами с разными размерами базовых типов. Например, если вы записываете целочисленное число в двоичном виде в файл, используйте тип фиксированного размера, например, `std::uint32_t`, но не `int`, `long` или `std::size_t`, которые могут иметь разные размеры на разных платформах. Проблемой `little-endian/big-endian` можно пренебречь.
- Для работы с байтами рекомендуется использовать тип `std::uint8_t` или по крайней мере `unsigned char`, но не `char`, т.к. с ним могут быть проблемы со знаковостью.
- Протестируйте решение на типичных граничных случаях, удостоверьтесь, что после разжатия сжатого файла получается исходный файл:
 - для пустого файла;
 - для файла, содержащего ровно один байт;
 - для файла, содержащего байт с кодом ноль;
 - для файла, содержащего байт с кодом 255;
 - для файла, содержащего ровно 256 различных байт;
 - для файла, содержащего только два одинаковых байта;
 - для больших текстовых файлов;
 - для больших бинарных файлов;
 - для больших файлов, содержащих случайные байты и, соответственно, плохо сжимающиеся;

Используйте прилагающийся к условию архив с тестами.

- В случае, если выходной файл существует, его необходимо перезаписать. Для реализации такого поведения вам не требуется проверять существование файла: перезаписать файл, если он существует, это поведение по умолчанию при открытии файла на запись.

Формат сдачи. В директории `ha2` в репозитории должны быть четыре файла: `main.cpp`, `huffman.hpp`, `huffman.cpp` и `Makefile` или `CMakeLists.txt`, если вы используете CMake.

Таким образом, ваша директория в Subversion должна выглядеть следующим образом:

```
ha2
├─ Makefile
├─ main.cpp
├─ huffman.cpp
└─ huffman.hpp
```

Сроки сдачи. Будет три срока, к которым можно будет сдавать домашнее задание:

- 8:00 8 ноября (воскресенье),
- 8:00 15 ноября (воскресенье),
- 8:00 22 ноября (воскресенье).

Если домашнее задание не принимается с первой попытки, его можно попробовать сдать со следующей попытки.