

Задачи

Решение задачи можно присылать (предпочтительнее ссылка на github, и т.п.) на vkonoovodov@gmail.com. Принимается не более одного решения от каждого слушателя.

Необходимо написать программу, которая выполняет сериализацию/десериализацию графа.

- В режиме `./run -s -i input.txt -o graph.bin` программа читает граф из файла `input.txt` в описанном ниже формате и записывает файл `graph.bin` в бинарном формате (тут нет ограничений).
- В режиме `./run -d -i graph.bin -o output.txt` программа должна считывать `graph.bin` и записывать его в том же текстовом формате в `output.txt`.
- Получившийся `output.txt` может отличаться от `input.txt` только перестановкой строк.

Основная задача — минимизировать размер `graph.bin`, не потеряв исходные данные.

Формат входных данных. Рассматриваются неориентированные взвешенные по рёбрам графы без кратных рёбер, с возможными петлями, без изолированных вершин. Вершины графа имеют целочисленные идентификаторы ($\text{id} \in [0, 2^{32} - 1]$). В каждой строке входного файла описывается одно ребро в виде трёх разделённых по символу табуляции целых чисел — идентификаторы смежных вершин и вес ребра (в таком порядке). Вес каждого ребра лежит в отрезке $[0, 2^8 - 1]$.

Если среди описанных рёбер не упоминается некоторая вершина с идентификатором v , можно считать, что такой вершины нет.

Как будет устроена проверка.

- Присланный код должен компилироваться без сторонних библиотек, не должен использовать `gzip`, `boost`, `proto`, и т.п.
- На вход в файле `input.txt` полученной программе будет подан граф из $N \approx 10^6$ вершин (≤ 100 MB), заведомо обладающий следующими свойствами:
 - (1) Граф разреженный (число рёбер не превосходит $5N$).
 - (2) Граф имеет близкое к степенному распределение степеней вершин. То есть существуют некоторые константы $\gamma > 1$ и c , для которых доля вершин степени d аппроксимируется при большей части значений d оценкой $\frac{c}{d^\gamma}$.
- Полученный файл `graph.bin` будет независимо десериализован той же программой в другом режиме в `output.txt`.
- Множества строк файлов `output.txt` и `input.txt` должны совпадать. В случае, если это не так, решение не принимается.
- Время работы программы не должно превосходить 30 минут.
- Баллы за задание будут выставлены после окончания принятия всех решений по размеру файла `graph.bin`.