

Преобразование Барроуза Уилера

Студент 413 группы
Гребенников Никита

Преобразование Барроуза Уилера (Burrows-Wheeler transform, BWT)- это алгоритм, используемый в техниках сжатия данных для преобразования исходных данных. BWT используется в архиваторе bzip2.

Краткое описание и решаемые задачи

Меняет порядок символов во входной строке таким образом, что повторяющиеся подстроки образуют на выходе идущие подряд последовательности одинаковых символов. Таким образом BWT выполняет задачу сжатия исключением повторяющихся подстрок.

Описание алгоритма

Когда символьная строка трансформируется при помощи BWT, ни один из её символов не изменяется. Оно просто меняет порядок символов. Если в исходной строке есть подстроки, которые встречаются часто, тогда трансформированная строка будет иметь некоторые места, где одиночный символ повторяется несколько раз подряд. Это полезно для компрессии, так как ведёт к облегчению сжатия строки, которая состоит из повторяющихся символов, при помощи таких техник, как кодирование длин серий.

Например, строка:

SIX.MIXED.PIXIES.SIFT.SIXTY.PIXIE.DUST.BOXES

трансформируется в эту* строку, которая легче сжимается, потому что содержит много повторяющихся символов:

TEXYDST.E.XIIXIXSMPPSS.B...S.EEUSFXDIOIIIT

bzip2

-утилита для сжатия данных, реализация алгоритма Барроуза Уилера.

Единовременно может выполнять только одну операцию: либо сжатие, либо распаковку и только для одного файла. При сжатии bzip2 добавляет к имени файла расширение «.bz2»

-сжимает большинство файлов эффективнее, но медленнее, чем более традиционные утилиты gzip или zip. В этом отношении он похож на другие современные алгоритмы сжатия.

-выполняет сжатие данных с существенной нагрузкой на процессор применяют, если нет ограничений на время сжатия и на нагрузку на процессор, например, для разовой упаковки большого объёма данных.

-в некоторых случаях bzip2 уступает по эффективности сжатия архиваторам 7-Zip и rar, метод сжатия bzip2 уступает по эффективности сжатия на 10-15 %, но при этом в 2 раза быстрее при сжатии и в 6 раз быстрее при распаковке.