Курсовая работа по курсу дискретного анализа: diff

Выполнил студет группы М8О-301Б-21: Кирилин Иван Олегович
Условие задачи:

Даны две строки со словами. Необходимо найти наибольшую общую подпоследовательность слов. Обратите внимание на ограничения по памяти, стандартное решение при помощи дп не подойдет. Попробуйте использовать ассимптотически линейное по памяти решение

Формат ввода

Две строки со словами разделенными пробелом.

В каждой строке не более 104 слов. Суммарная длина всех слов не превышает 2·10^5.

Формат вывода

В первой строке выведите одно число L — длину наибольшей общей подпоследовательности. Во второй строке через пробел выведите L слов — найденную подпоследовательность.

Подход к решению:

Для решения задачи можн воспользоваться алгоритмом Хиршберга, алгоритм основан на методе “Разделяй и влавствуй”. Он разделяет задачу на подзадачи, в котором выбирается точка разделения, точка разделения X есть середина строки(mid). Точка разделения Y (j) выбирается такой, что:
$|LCS(x[0, mid], y[0, j]) + LCS(x2.reverse(), y[j+1, n].reverse)| \rightarrow max$
После получения LCS для каждой из подзадач происходит комбинирование в одну общую LCS. Базовый случай – когда в X остается 1 элемент, если он есть в Y, то записываем его в LCS, иначе – нет.

Ассиптотическая сложность времени работы алгоритма: O(nm)
Пространственная сложность алгоритма: O(n+m)

Описание программы:

В программе содержится, как основная функция Hirschberg_LCS, которая выоплняет разделение, так и вспомогательная функция LCS, которая возвращает длину последнюю строку обычно LCS (реализованным ДП), вместе с ее длиной.

Тесты производительности

	Base case (as example)	Small length, 4,0K (both strings)	Medium length, 24K (both strings)	Huge length, 1M (both strings)
Time consumption	real 0m0,006s	0m0,008s	0m0,498s	0m44,358s
Space usage	87,765 bytes allocated	333,900 bytes allocated	8,205,541 bytes allocated	37,643,925 bytes allocated

Прим. Чтобы имерить время использовалась утилита time, проверка на память – valgrind

Недочеты

На огромных файлах время работы все еще достаточно большое

Выводы

Алгоритм очень полезен для ускорения некотрых сравнительных утилит(пр. Diff), он сокращает количетво потребляемой памяти, что делает уитлиту более универсальной. Сам курсовой проект мне понравился, так как интересна область алгоритмов, в том числе и рекурсивных и усовершенствованных