# Преобразование Барроуза-Уиллера

5 января 2023 г. 1:27

Пусть имеется последовательность на выходе источника.
Берем все возможные циклические сдвиги (т.е. берем последнюю букву и ставим в начало и т.д.), сортируем их в лексикографическом порядке и формируем из них таблицу.
Результатом преобразования являются:

1. Последний столбец таблицы
2. Номер строки, которая соответствует исходной последовательности

| | | | |
|---|---|---|---|
| 0 | A | NIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_ | C |
| 1 | A | NNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_ | C |
| 2 | A | S_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO | _ |
| 3 | A | S_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO | _ |
| 4 | C | ANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE | _ |
| 5 | C | ANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE | _ |
| 6 | D | O_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD | _ |
| 7 | D | O_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT | _ |
| 8 | D | _DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOU | L |
| 9 | D | _WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOU | L |
| 10 | E | _CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_ | W |
| 11 | E | _CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_ | W |
| 12 | E | _SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_ | W |
| 13 | E | _WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_ | W |
| 14 | F | _WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CAN | I* |
| 15 | H | OULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_ | S |
| 16 | I | F_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CA | N |
| 17 | L | D_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHO | U |
| 18 | L | D_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WO | U |
| 19 | N | IF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_C | A |
| 20 | N | NOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_C | A |
| 21 | N | OT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CA | N |
| 22 | O | T_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CAN | N |
| 23 | O | ULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_S | H |
| 24 | O | ULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_ | W |
| 25 | O | _AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_ | D |
| 26 | O | _AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_ | D |
| 27 | S | HOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE | _ |
| 28 | S | _WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_ | A |
| 29 | S | _WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_ | A |
| 30 | T | _DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANN | O |
| 31 | U | LD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SH | O |
| 32 | U | LD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_W | O |
| 33 | W | E_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS | _ |
| 34 | W | E_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF | _ |
| 35 | W | E_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD | _ |
| 36 | W | E_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS | _ |
| 37 | W | OULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE | _ |
| 38 | _ | AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_D | O |
| 38 | _ | AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_D | O |
| 40 | _ | CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_W | E |
| 41 | _ | CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_W | E |
| 42 | _ | DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOUL | D |
| 43 | _ | DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNO | T |
| 44 | _ | SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_W | E |
| 45 | _ | WE_CANIF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_A | S |
| 46 | _ | WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CANI | F |
| 47 | _ | WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_WE_WOUL | D |
| 48 | _ | WE_WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_A | S |
| 49 | _ | WOULD_WE_SHOULD_DO_AS_WE_CANIF_WE_CANNOT_DO_AS_W | E |

После формирования таблицы мы получили номер строки (16) и последний столбец:
CC_____LLWWWWISNUUAANNHWDD_AAOOO_____OOEEDTESFDSE
Видно, что сформировалась последовательность с явными повторениями символов - это можно учесть при сжатии.
В итоге передается этот последний столбец и номер строки, содержащий исходную последовательность

Как декодировать этот последний столбец?
Если отсортировать его в алфавитном порядке, то получим первый столбец.
Таким образом, известен номер строки (16), первый и последний столбец таблицы.
Надо найти второй столбец.
Смотрим строку 16, в первом столбце видим I.
Ищем I в последнем столбце.
Нашли строку, снова смотрим первый её столбец.
И т.д.
Но в последнем столбце присутствуют одинаковые символы - какой выбрать?
Считаем, какой по счету этот символ в последнем столбце, в первом столбце ищем такой же по счету этот символ.

---

Кодирование последнего столбца "стопкой книг"

- CC_____LLWWWWISNUUAANNHWDD_AAOOO_____OOEEDTESFDSE
- Будем писать *esc* каждый раз, когда появляется новый символ и число различных букв между предыдущим и текущим появлением буквы.
- Получим: *esc* 0 *esc* 0 0 0 0 0 *esc* 0 *esc* 0 0 0 *esc esc esc esc* 0 *esc* 0 2 0 *esc* 6 *esc* 0 9 5 0 *esc* 0 0 2 0 0 0 0 1 0 *esc* 0 4 *esc* 2 10 *esc* 4 2 3
- Применяем нумерационное кодирование:
  - Кодирование композиции: 35 бит.
  - Кодирование последовательности: 94 бита.

- Передача символов, соответствующих *esc*: $8 \times 15 = 120$ бит.
- Передача индекса исходной строки в преобразовании: 6 бит.

Итого: 255 бит.