

Преобразование Барроуза-Уилера

Алексей Владыкин

СПбГУ ИТМО

20 мая 2011

- Michael Burrows (1963–)
- David Wheeler (1927–2004)
- Burrows–Wheeler transform, 1984 г.
- Преобразование само по себе не является сжатием, но повышает сжимаемость данных алгоритмами типа move-to-front или RLE.
- Используется в BZIP2 (BWT+MTF+Huffman)

Прямое преобразование

- Взять блок входных данных длиной n , дописать в конец специальный символ $\$$.
- Сгенерировать все циклические сдвиги этого блока и занести их в строки таблицы $(n + 1) \times (n + 1)$.
- Отсортировать строки таблицы.
- Взять последний столбец таблицы.

Прямое преобразование

- Взять блок входных данных длиной n , дописать в конец специальный символ $\$$.
- Сгенерировать все циклические сдвиги этого блока и занести их в строки таблицы $(n + 1) \times (n + 1)$.
- Отсортировать строки таблицы.
- Взять последний столбец таблицы.

- На практике: циклический сдвиг однозначно задается размером сдвига, поэтому хранить все сдвинутые строки не надо.

Обратное преобразование

- Создать пустую таблицу $(n + 1) \times (n + 1)$ и повторять $n + 1$ раз:
- — вставить блок первым столбцом таблицы;
- — осортировать строки таблицы.
- Взять строку таблицы, заканчивающуюся на \$.


Обратное преобразование


- Создать пустую таблицу $(n + 1) \times (n + 1)$ и повторять $n + 1$ раз:
 - — вставить блок первым столбцом таблицы;
 - — осортировать строки таблицы.
- Взять строку таблицы, заканчивающуюся на \$.
- На практике: нет необходимости хранить таблицу и сортировать строки. Декодирование делается по первому и последнему столбцу.

После преобразования

- move-to-front
- run-length encoding
- zero running
- Huffman

Рекомендуемая литература

 Новиков Ф. А.
Дискретная математика для программистов.
СПб.: Питер, 2000. — 304 с.: ил. // Глава 6

 Романовский И. В.
Дискретный анализ. — 3-е изд., перераб. и доп.
СПб: Невский Диалект; БХВ Петербург, 2003. — 320 с.: ил.
// Глава 5