

Описание структуры закодированного файла Huffman Codec

Структура закодированного файла

Закодированный файл состоит из двух основных частей: заголовка (*header*) и непосредственно закодированных данных (*compressed data*). Каждая часть имеет строгий формат, описанный ниже.

1. Заголовок (*Header*)

Заголовок содержит метаданные о файле и словарь восстановления (*recovery map*). Он имеет следующий формат:

```
HUFFMANHEADER
<Original File Name>
<Dictionary Size>
<Dictionary Entries>
```

- **HUFFMANHEADER:** Строка-маркер, указывающая, что файл является архивом Хаффмана.
- **Original File Name:** Имя исходного файла, используемое для восстановления.
- **Dictionary Size:** Количество записей в словаре восстановления.
- **Dictionary Entries:** Словарь восстановления, где каждая запись представлена как символ и соответствующий ему код. Формат записи:
 - `<symbol><code>` для символов, отличных от `'\n'`.
 - `\n<code>` для символа новой строки.

2. Закодированные данные (*Compressed Data*)

После заголовка расположены закодированные данные в следующем формате:

```
<Padding Info><Encoded Data>
```

- **Padding Info:** 8 бит, указывающие, сколько нулевых битов было добавлено в конец закодированных данных для выравнивания длины до байта.
- **Encoded Data:** Последовательность битов, представляющая закодированное содержимое исходного файла.

Пример структуры

Пример для файла `example.txt` с содержимым "HELLO":

```
HUFFMANHEADER
example.txt
4
H00
E01
L10
011
00000010 01000101 10011001 10000011
```

- Заголовок:
 - HUFFMANHEADER
 - Имя файла: `example.txt`
 - Размер словаря: 4
 - Словарь: H00, E01, L10, 011
- Закодированные данные:
 - Padding Info: 00000010 (2 дополнительных нуля).
 - Encoded Data: 01000101 10011001 10000011