

## ШИФРОВАНИЕ ХАФФМАНОМ

В работе рассматривается программная реализация алгоритма Хаффмана и его применение для сжатия и распаковки данных.

### ВВЕДЕНИЕ

Коды Хаффмана – оптимальный префиксный код, используемый в сжатии данных без потерь. Коды Хаффмана являются результатом работы алгоритма Хаффмана, разработанного Дэвидом Хаффманом в 1952 году.

### I. ПРИНЦИП РАБОТЫ

Все символы обычного текстового файла представляются одним (ASCII) или двумя (Unicode) байтами. В обеих кодировках каждый символ кодируется одинаковым количеством битов.

Символ	Двоичное представление	Десятичное представление
A	01000001	65
B	01000010	66
C	01000011	67
...	...	...
X	01011000	88
Y	01011001	89
Z	01011010	90

Главный принцип алгоритма Хаффмана состоит в создании специальных кодов, назначаемых каждому из символов сжимаемого сообщения таким образом, что наиболее часто встречаемому символу назначается самый короткий код Хаффмана.

Таким образом, часто встречаемые в сообщении символы будут уже представлены не восьмью, но уже меньшим числом бит, тем самым уменьшая размер всего сообщения.

Другое важное свойство кодов Хаффмана – это то, что ни один код не является префиксом другого кода. Делается это с целью избежать неопределенности при чтении потока данных.

Алгоритм Хаффмана можно разбить на несколько простых этапов:

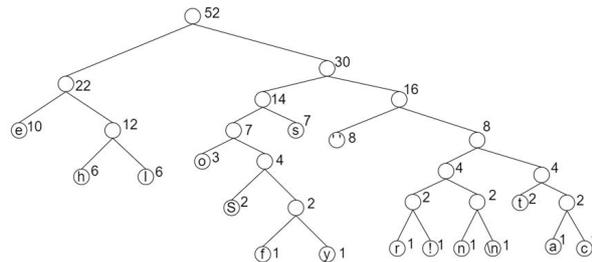
- Создание частотной таблицы символов сообщения.

*Плотников Владислав Вадимович*, студент 1 курса факультета компьютерных систем и сетей Белорусского государственного университета информатики и радиоэлектроники, vadimovich74@gmail.com

*Кресс Владислав Дмитриевич*, студент 2 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, vladkress@yandex.by

*Научный руководитель: Шатилова Ольга Олеговна*, старший преподаватель кафедры ВМиП, shatilova@bsuir.by

- Построение дерева Хаффмана на базе частотной таблицы символов.
  - Обход дерева Хаффмана и присваивание соответствующего кода Хаффмана каждому символу исходного сообщения.
- Для сообщения «She sells seashells on the shore of the Seychelles!» дерево Хаффмана будет выглядеть следующим образом:



Как только дерево построено, можно приступить к шифрованию сообщения. Движемся от начала корня ко всем листьям, запоминая «маршрут»: переход в левое поддерево – ноль, а в правое – единица. То есть для символа *e* код будет 00, а для символа *h* – 010. Как можно заметить, самые встречаемые символы находятся ближе к корню дерева, а значит «маршрут» до них (т.е. код Хаффмана) будет самым коротким.

### II. РЕЗУЛЬТАТЫ РАБОТЫ

- Был изучен алгоритм Хаффмана, его свойства, достоинства и недостатки.
- Было реализовано программное средство сжатия и распаковки данных методом Хаффмана на языке Java.

### Список литературы

1. Роберт Лафоре «Структуры данных и алгоритмы Java».
2. Khalid Sayood «Introduction to Data Compression».