


ХІІІ МЕЖДУНАРОДНАЯ МОЛОДЕЖНАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ С
ЭЛЕМЕНТАМИ НАУЧНОЙ ШКОЛЫ "ПРИКЛАДНАЯ МАТЕМАТИКА И ФУНДАМЕНТАЛЬНАЯ
ИНФОРМАТИКА"



ИСПОЛЬЗОВАНИЕ АЛГОРИТМА СЖАТИЯ ДЛЯ ОПТИМИЗАЦИИ РАБОТЫ ГЕНЕРАТОРА ПРЕЗЕНТАЦИЙ

Подготовил: Студент Кудряшов Е.А.



Цель исследования

Исследование заключается в применении одного из ключевых аспектов теории информации – методов сжатия данных.



Цифровая образовательная платформа



Методы сжатия данных

Методы сжатия данных с точки зрения возможности использования их для сжатия информации любого типа можно разделить на универсальные и специальные.

Далее идут универсальные обратимые методы сжатия информации. Здесь можно выделить три основных группы методов сжатия.

- Методы преобразования блока;
- Преобразование потока;
- Статистические методы сжатия.



Статистические методы сжатия

Статистические методы можно разделить на две группы методов.

- Адаптивные или поточные методы;
- Блочные методы.



Блочные методы

В этой группе методов отдельно кодируется статистика блока и добавляется к сжатому блоку.

Примеры блочных методов:

статические варианты методов Хаффмана, Шеннона-Фано, арифметического кодирования для потоков элементов.



Критерии оценки методов сжатия

Основными свойствами какого-либо алгоритма сжатия данных являются:

- **качество (коэффициент или степень) сжатия**, т. е. отношение длины (в битах) сжатого представления данных к длине исходного представления;
- **скорость кодирования и декодирования**, определяемые временем, затрачиваемым на кодирование и декодирование данных;
- **объем требуемой памяти.**



Основные формулы

Коэффициент сжатия: $K_{compress} = \frac{S_{output}}{S_{input}}$

Скорость сжатия R , единицей измерения скорости сжатия является «количество кодовых бит, приходящихся на отсчет данных источника»

Скорость сжатия: $R = \frac{k}{n}$

Оптимальную длину кода можно определить через частоту (вероятность) появления элемента с использованием понятия энтропия символа.

Энтропией символа: $H_c = p_c \cdot \log_2 p_c$



Алгоритм Хаффмана

Идея алгоритма Хаффмана состоит в том, чтобы наиболее часто встречающиеся символы имели более короткие коды, символы, встречающиеся реже всего, имели очень длинный код. Алгоритм сжатия данных Хаффмана, как канонический, так и его адаптивные версии, обладают достаточно высокой эффективностью и лежат в основе многих других методов, используемых в алгоритмах сжатия данных.

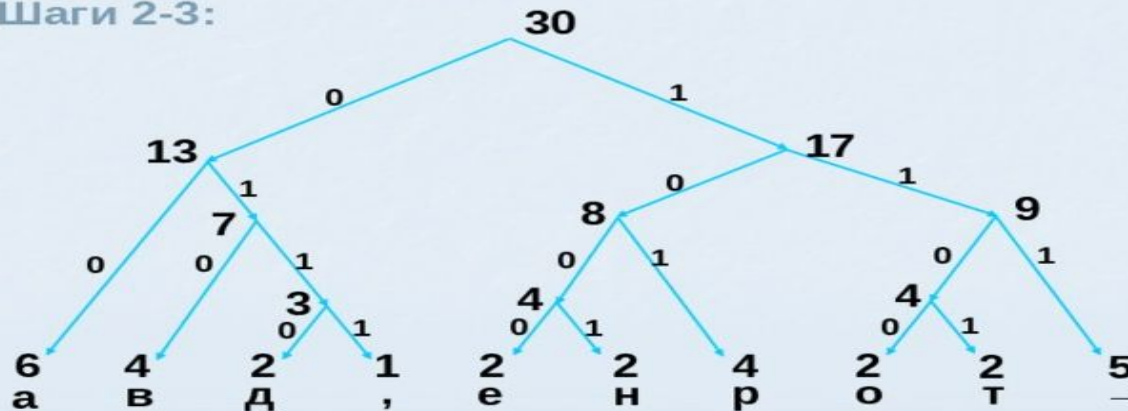
Алгоритм Хаффмана

Пример. Построить код Хаффмана для фразы:
на дворе трава, на траве дрова

Шаг 1:

а	в	д	,	е	н	р	о	т	_
6	4	2	1	2	2	4	2	2	5

Шаги 2-3:



Алгоритм Шеннона-Фано

Кодирование Шеннона-Фано было первым методом, разработанным для поиска хороших кодов переменного размера.

	Prob.	Steps				Final
1.	0.25	1	1			:11
2.	0.20	1	0			:10
3.	0.15	0	1	1		:011
4.	0.15	0	1	0		:010
5.	0.10	0	0	1		:001
6.	0.10	0	0	0	1	:0001
7.	0.05	0	0	0	0	:0000



Результаты тестового исследования

	Размер входного файла	Формат данных	Размер выходного файла	Коэффициент сжатия
Хаффмана	155 кб	JSON	98 кб	0.63
Шеннона-Фано	155 кб	JSON	122 кб	0.78



Итоги

Подводя итоги, оба алгоритма имеют свои преимущества и недостатки в зависимости от конкретной ситуации. Алгоритм Хаффмана может быть более ресурсоемким при построении дерева, но может обеспечить лучшую степень сжатия. Алгоритм Шеннона-Фано может потребовать меньше памяти, но может не обеспечивать такое же высокое сжатие. В итоге выбор алгоритма будет зависеть от конкретных требований по ресурсам и степени сжатия в конкретной задаче.



Библиографический список

- 1.Сэломон Д. Сжатие данных, изображения и звука. — М.: Техносфера, 2004. — С. 368.
- 2.Mark Nelson и Jean-Loup Gailly Introduction to Data Compression. Khalid Sayood University of Nebraska. 702 p.
- 3.Ватолин Д., Ратушняк А., Смирнов М., Юкин В.В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. - М.: ДИАЛОГ-МИФИ, 2003. - 384 с.
- 4.David Salomon Data Compression: The Complete Reference. .California State University, Northridge Northridge, CA 91330-8281. 920 p.