



**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

## **Классификация методов подсчета информационной энтропии**

**Студент: Хамзина Регина Ренатовна ИУ7-73Б  
Руководитель: Оленев Антон Александрович**

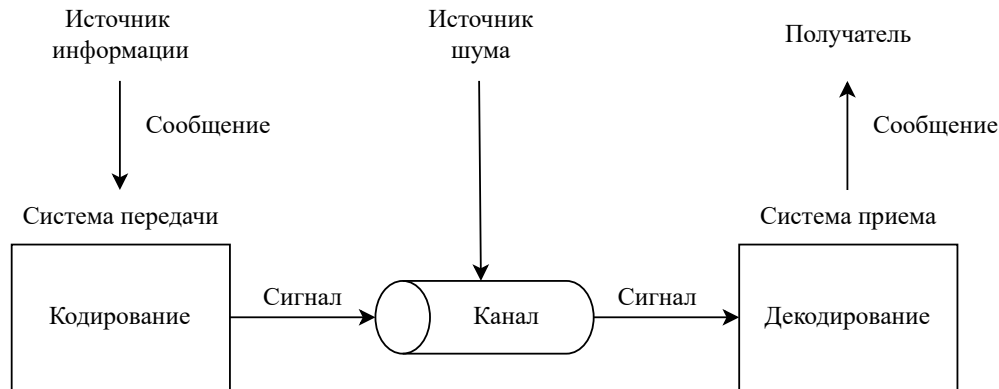
# Цель и задачи

**Цель:** классификация методов подсчета информационной энтропии.

**Задачи:**

- провести анализ предметной области: рассмотреть основные определения, изучить свойства информационной энтропии и ее связь со сжатием данных;
- описать существующие методы подсчета информационной энтропии;
- выделить критерии сравнения описанных методов;
- провести сравнение методов по выделенным критериям.

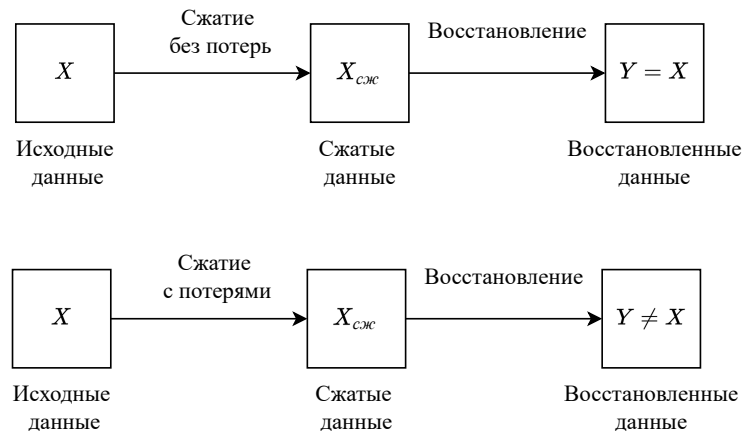
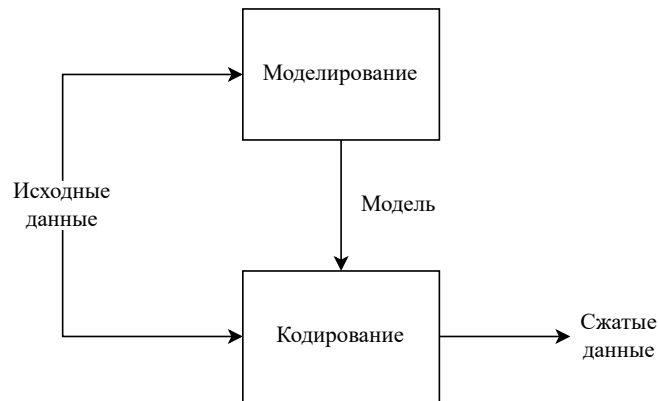
# Передача информации



$$H(X) = - \sum_{i=1}^n (p_i \cdot \log_a p_i), \quad (1)$$

где  $n \in \mathbb{N}$ ,  $p_i = P(X \sim x_i)$ ,  $\sum_{i=1}^n p_i = 1$ ,  $a > 1$ .

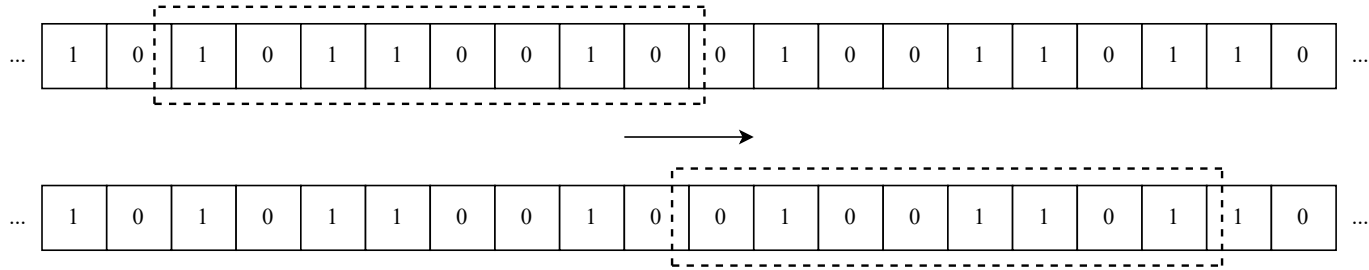
# Связь энтропии и сжатия данных



$$K_{сж} = \frac{L_{исх}}{L_{сж}}, \quad (2)$$

где  $L_{исх}$  — объем исходных данных  $X$ ,  $L_{сж}$  — объем сжатых данных  $X_{сж}$ .

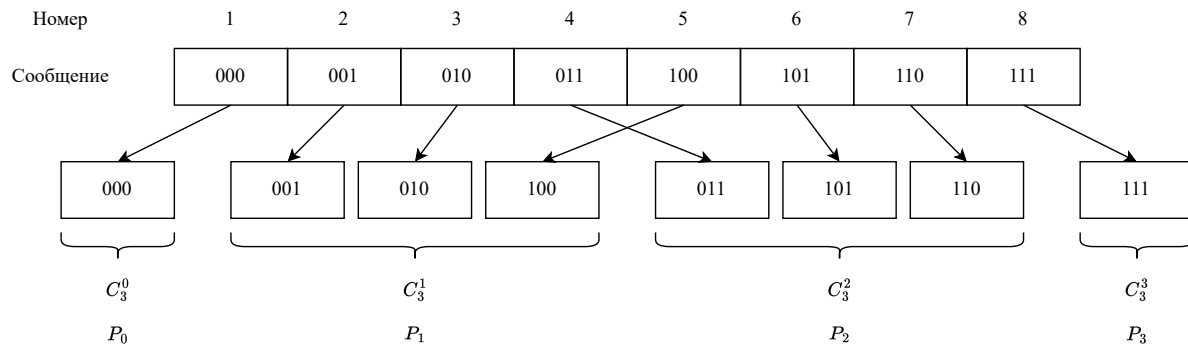
# Метод скользящего окна



$$H(X) = - \sum_{i=0}^{255} (p_i \cdot \log_2 p_i), \quad (3)$$

где  $p_i = \frac{k_i}{N}$  — вероятность появления байта в массиве байтов размером  $N$ ,  $k_i$  — число вхождений байта в массив байтов.

# Биномиальный метод



$$H(X) = - \sum_{k=0}^n (C_n^k \cdot P_k \cdot \log_2 P_k), \quad (4)$$

где  $P_k = p^k \cdot (1 - p)^{(n-k)}$ ,  $C_n^k = \frac{n!}{k! \cdot (n-k)!}$ .

# Сравнение методов

Метод	К1	К2	К3	К4
Скользящего окна	$O(N + 2^n)$	-	+	$2^n$
Биномиальный	$O(N + n)$	+	+	$2 \cdot (n + 1)$

- К1 — временная сложность;
- К2 — необходимость вычисления факториала;
- К3 — возможность распараллеливания вычислений;
- К4 — объем требуемой дополнительной памяти.

# Заключение

Цель научно-исследовательской работы достигнута.

При написании данной работы:

- проведен анализ предметной области: рассмотрены основные определения, изучены свойства информационной энтропии и ее связь со сжатием данных;
- описаны существующие методы подсчета информационной энтропии;
- выделены критерии сравнения описанных методов;
- проведено сравнение методов по выделенным критериям.