**Lab 7**

**Fano coding algorithm**

**Objective of the work** : To study the operating principle of the Shannon- Fano algorithm used for lossless data compression, to master the process of encoding and decoding symbols based on the probabilities of their occurrence, and to develop an algorithm for encoding and decoding a message.

**Theoretical part**

**The Shannon- Fano algorithm** is an entropy coding method that assigns variable-length codes to symbols based on their probability of occurrence. Symbols that occur more frequently are assigned shorter codes, while less likely symbols are assigned longer codes. This allows for a reduction in the overall amount of data required to transmit or store information.

**The principle of the algorithm:**

1. For each symbol, the probability of its appearance in the message is determined.
2. Symbols are sorted in descending order of probability.
3. The set of symbols is divided into two groups in such a way that the sum of the probabilities of each group is as equal as possible.
4. The left group is assigned the bit "0", the right one - "1".
5. The division process continues for each group until each symbol is assigned a unique code.

**Practical part**

**Objectives of the laboratory work:**

1. Fano codes for the symbols of the original message.
2. Implement the Shannon- Fano algorithm to encode a string in a programming language of your choice (e.g. Python).
3. Evaluate the compression efficiency based on the generated codes.

***p.s. Before defending your work, manually calculate the code for the string " ABCABEAAFB " and compare the result with the software calculation.***

**Progress of work:**

1. **Determination of symbol frequencies:**

* Enter the string to encode. For example: "ABABACD".
* Count how many times each character appears in the string.
* Calculate the probability of each symbol (probability = symbol frequency / total number of symbols).

2. Construction of Shannon- **Fano code :**

* Sort the symbols in descending order of their probabilities.
* Divide the list of symbols into two groups with approximately equal sum of probabilities.
* Assign "0" to the characters in the first group and "1" to the characters in the second.
* Repeat the division process for each group until you have codes for all the symbols.

**Example:**

Message given: ABABACD

1. We calculate the frequencies of symbols (their number): A: 3 times, B: 2 times, C: 1 time, D: 1 time
2. We calculate the probabilities of symbols: A: 3/7, B: 2/7, C: 1/7, D: 1/7
3. We assign codes:
   * We sort the symbols by probabilities: A, B, C, D.
   * We divide into groups: {A}, {B, C, D}.
   * We assign: A — "0", {B, C, D} — "1".
   * In group {B, C, D}:
     + We divide {B}, {C, D}.
     + We assign B to "10", {C, D} to "11".
     + In the group {C, D}: C is "110", D is "111".

The following final codes were obtained: A: 0, B: 10, C: 110, D: 111