Программа работает по следующему алгоритму:

1. Считываются входные данные и строится кодовое дерево (одно дерево для простого варианта кода, и N деревьев для контекстного, где N – количество различных символов во входных данных)
2. На основе полученных кодовых слов производиться кодирование и запись входного файла, информация об используемых кодовых словах записывается в заголовок файла.
3. Производится чтение закодированных данных и их декодирование.

В качестве тестового файла использовался файл, сгенерированный с помощью марковской цепи, описываемой следующей матрицей переходных вероятностей:

[0.743278 0.0278329 0.0559099 0.131626 0.0413525;

0.847072 0.0146794 0.126469 0.000161 0.0117801;

0.0946684 0.258797 0.232765 0.14655 0.26722;

0.125095 0.0801111 0.0470901 0.0587176 0.688986;

0.336406 0.398602 0.259163 0.00433363 0.00149548];

Энтропия данного источника равна 1.38 бит/символ.

Исходный размер файла - 200(байт)

Размер файла при энтропийном кодировании 34,5(байт)

Размер файла сжатого с помощью кода Хаффмана - 60(байт)

Размер файла сжатого с помощью контекстного кода Хаффмана - 50(байт)

При увеличении размеров файла различия между теоретическими и практическими значениями вероятностей будут стремиться к нулю, однако при кодировании файлов маленького размера эффективнее использовать вероятности, полученные при анализе входных данных.