

СЖАТИЕ ИНФОРМАЦИИ НА ОСНОВЕ НЕРАВНОМЕРНЫХ БИНОМИАЛЬНЫХ КОДОВ

студ. Харченко С.Н., студ. Сурженко А.И.,
ст. преп. Протасова Т.А.

Современное мировое сообщество, перейдя на новый этап своего информационного развития встретило ряд проблем. Одной из наиболее важной из них стала проблема накопления, распространения и хранения получаемых объёмов информации. Ежедневно создаётся документации, имеющей инновационный характер, столько, что она занимает большую комнату. Для хранения возрастающих с каждым днём объёмов информации человечество уже давно ушло от бумаги, и пришло к магнитным и оптическим накопителям данных. Но и это не решило поставленную проблему. Результатом чего стали интенсивные разработки в области создания различных методов сжатия информации.

Как известно, сжатие сводится к обработке информации таким образом, чтобы получить выигрыш в размере выходных данных по сравнению со входными. При использовании биномиального кода в качестве неравномерного нет возможности построения кодового дерева для нахождения оптимальных кодовых комбинаций. Поэтому приходится пользоваться только вероятностями появления того или иного символа на входе и заранее сгенерированной таблицей для упаковки. Для того чтобы получать большие коэффициенты сжатия был разработан алгоритм анализа статистических данных. Информация подаётся на вход в виде больших пакетов произвольной длины. После получения данные представляются в виде последовательности байт и производится вычисление вероятностей встречающихся символов, на основе которых

выбираются оптимальные параметры биномиального кода для получения наилучшего коэффициента сжатия: в зависимости от количества различных символов, от разброса вероятностей, а также кодового расстояния между комбинациями. Далее, при помощи счётчика с заданными параметрами, непосредственно генерируются комбинации, которые сортируются по длине. Каждому байту ставится в соответствие уникальная последовательность бит, выраженная в биномиальном коде. На выходе получаем данные, которые снабжаются заголовком, несущим в себе сгенерированную таблицу соответствий (которая практически уникальна для каждого обрабатываемого пакета). Иногда дополнительно формируется контрольная сумма для сигнализации о возможных ошибках. При анализе такого пакета считывается таблица, по которой определяются параметры кода, являющиеся критерием в определении конца комбинации. Далее данные декодируются с помощью этой таблицы.

Использование данного метода позволяет значительно увеличить средний коэффициент сжатия за счёт того, что для каждого нового пакета, на основе статистических данных о нём, генерируется новая таблица оптимизированного неравномерного кода, дающего максимальное сжатие только для данного блока.

Перспективным направлением исследований является разработка эффективного метода сжатия непосредственно текстовой информации. Особенно в случае, если решающим является не скорость; а получаемый коэффициент сжатия.

Программный эквивалент предложенного метода сжатия на основе неравномерных биномиальных кодов отличается простотой реализации и исследований при дальнейшем построении устройства.