

Е. В. СИЛКОВ

Научный руководитель А. И. ЯКИМОВ, канд. техн. наук, доц.

ГУ ВПО «Белорусско-Российский университет»

Одним из способов ускорения передачи данных является их сжатие, т. к. информация в сжатом виде требует меньше объема памяти. Для сжатия данных используются методы (алгоритмы) сжатия данных, при использовании которых учитывается следующие основные параметры: коэффициент компрессии (сжатия), скорость компрессии (сжатия), скорость декомпрессии (распаковки). Все эти параметры зависят от применяемого алгоритма сжатия данных, на выбор которого влияет тип сжимаемой информации (текстовая информация или графическая, аудио или видео), а также сама конструкция этой информации (это может быть произвольный текст, данные из баз данных, код на каком-либо языке программирования и т. д.). Еще две важные характеристики алгоритма сжатия – объемы памяти, необходимые для компрессии и для декомпрессии (для хранения данных, создаваемых и/или используемых алгоритмом).

Существуют базовые алгоритмы сжатия информации: метод Шеннона-Фано, метод Хаффмана и его адаптивная версия, методы Лемпеля-Зива, арифметическое кодирование, интервальное кодирование и др.

При разработке технологии повышения скорости данных были улучшены методы сжатия текстовых данных (алгоритм Хаффмана и алгоритм Шеннона-Фано) путем кодирования не только символов, но и часто встречающихся слов.

Исследование показало, что при кодировании лишь одних символов методом Шеннона-Фано коэффициент сжатия составил 1,652; методом Хаффмана – 1,663. При кодировании символов и первых 15 самых встречающихся слов методом Шеннона-Фано коэффициент сжатия увеличился на 0,026 (+2,6 %); методом Хаффмана – на 0,022 (+2,2 %). При кодировании символов и первых 50 наиболее часто встречающихся слов методом Шеннона-Фано коэффициент сжатия увеличился на 0,039 (+3,9 %); методом Хаффмана – также на 0,039 (+3,9 %). Однако на это ушло гораздо больше времени: длительность сжатия увеличилась более чем в 100 раз.

Исследованная модификация сжатия со словами эффективна, если на сервере или любом другом ресурсе уже хранятся текстовые данные (или если необходимо сократить объем передаваемых данных, но временем сжатия можно пренебречь).