

УДК 004.627

МНОГОПОЛЬЗОВАТЕЛЬСКАЯ СИСТЕМА СЖАТИЯ ДАННЫХ С ОБЩИМ СЛОВАРЕМ

*Ю.А. Зубань, доцент, к. т. н., СумГУ, yuzha@ukr.net;
В.В. Петров, инженер, институт прикладной физики
АН Украины*

На сегодняшний день существует много подходов к сжатию информации и алгоритмов, их реализующих. Наиболее распространенными являются словарные методы. Наиболее эффективные из них (семейство алгоритмов LZ78) основываются на замене строки символов в сжимаемом потоке на соответствующий номер элемента в словаре. Эффективность сжатия непосредственно зависит от размера словаря и его содержимого. Словарь, как правило, создается по мере просмотра сжимаемого потока символов, поэтому начальные последовательности символов сжимаются менее эффективно, чем последние, когда словарь сформирован и в него входит достаточное количество записей, повторяющихся в сжимаемом потоке.

Эффективность словарных методов сжатия можно существенно повысить, если начинать процесс сжатия с готовым словарем, учитывающим наиболее вероятные последовательности символов в сжимаемом потоке, например слова украинского, английского или русского алфавита. По мере работы со словарем его содержимое может пополняться новыми словами и фразами.

Совместное использование словаря многими пользователями позволит динамически корректировать информацию о часто употребляемых фразах, а также обеспечит доступ к сжатым данным только в пределах группы пользователей, имеющих доступ к словарю. Это может быть востребовано в корпоративных информационных системах