# US-5109433-A Compressing and decompressing text files

(1) TECHNICAL FIELD  
(2) This invention is in the field of reducing the storage space required for a file and, more particularly, to a method of compressing and decompressing text files.  
(3) BACKGROUND OF THE INVENTION  
(4) Storing digital data in a minimum space is frequently desired. There are many methods of compressing data to minimize the memory required for storing the data. One technique termed "Huffman compression" replaces bytes in the data text with bits. In Huffman compression, the unique code of bits used for a byte is based on the frequency of that particular byte in relation to other bytes in the text. When the compressed text is accessed, it is decoded by a logic tree technique of sequentially stepping through the individual bits to locate the byte represented by the bit code. Compressing and decompressing files using the Huffman technique is well known. Huffman, David A., "A Method for the Construction of Minimum-Redundancy Codes," Proceedings of the I.R.E., September 1952. A second current method of compressing text is the replacing of runs of characters with a run flag, a character, and a run count.  
(5) Current digital data compressing techniques, including Huffman compression and others, provide a first level compression of the full-format text. Unfortunately, text compressed using Huffman and other compression techniques may still be prohibitively large for some applications. Further, present techniques for decompressing the compressed text are too slow to be used in some environments. For example, a user or program may require immediate access to a readable form of the text file and not be able to wait while decompression and assembly into a readable form occur. In such environments, computer memory and disk storage are sacrificed for speed in accessing full-format text rather than compressing the text.  
(6) SUMMARY OF THE INVENTION  
(7) It is therefore an object of this invention to provide a method of compressing text sequentially through repeated levels to minimize the storage space required.  
(8) It is another object of this invention to provide a method of rapidly decompressing text to its original, full-format form for display to a user.  
(9) It is a further object of the invention to store help files in a compressed data base.  
(10) These and other objects of the invention, as will be apparent herein, are accomplished by generating a compressed text file having phrases that are digitally represented by a unique code of bits. A full text file in standard digital format ready for display on a screen is provided. A first level compressed text file is generated from the full text file by replacing runs of identical characters with a code identifying the character and the run count. The first level compressed text file is then further compressed by replacing frequently occurring phrases in the first level of compressed text file with a key phrase flag and an index byte. A third level compressed text file is generated from the second level compressed text file by replacing bytes in the second level of compressed text file with a unique string of bits.  
(11) The full text file is sequentially and serially compressed through three levels. That is, the full text file is compressed using a first technique. The once compressed text file is further compressed using a second technique. The twice compressed text file is further compressed using a third technique. By sequentially and serially compressing the full text file, the space required for storing the file is significantly smaller than the memory required to store the full text file or for storing a once compressed file. The particular sequence selected for compressing the text achieves a minimum file size.  
(12) The file containing the compressed text also contains the tables and indices required to access and then decompress any selected portion of the text. The file includes a topic index sequentially identifying the location of the compressed text. Context strings and a context map are provided to facilitate access to the topic index. A key phrase table and Huffman decode tree are also stored in the compressed file.  
(13) Compressed text is decompressed in a single pass within the same file containing the compressed text. When the compressed text is accessed, decompression of each phrase is completed prior to continuing with the next phrase. The text is output immediately upon being decompressed. The compressed text is thus rapidly decompressed and provided to a user.