

Title (en)
METHOD AND APPARATUS FOR DATA COMPRESSION.

Title (de)
VERFAHREN UND VORRICHTUNG ZUM VERDICHTEN VON DATEN.

Title (fr)
PROCEDE ET APPAREIL DE COMPRESSION DE DONNEES.

Publication
EP 0160672 A1 19851113 (EN)

Application
EP 84903871 A 19841017

Priority
US 54328683 A 19831019

Abstract (en)
[origin: WO8501814A1] A method and apparatus for compressing alphanumeric data that is stored or transmitted in the form of digital codes. A dictionary is created which assigns each word of the alphanumeric text and the punctuation that follows it to a unique address or token of, illustratively, up to 16 bits (two bytes). Each word in the alphanumeric text is then replaced by the address that refers to that word in the dictionary. Because the dictionary can contain up to $2^{16} = 65,536$ entries, it is more than adequate for the storage of the words associated with almost any book. Because only two bytes of information are needed to address any one of these 65,000 words, replacement of each word of text with two bytes of address information reduces the average number of digits required to store the text by a factor of about three. Further reductions of 25% or more in the length of the compressed text can be achieved in most cases by representing the most frequency used words with tokens that are shorter than two bytes in length. The number of bytes required to store the dictionary can be substantially reduced by storing the words in alphabetical order and taking advantage of the redundancy in characters that results. Thus, if the second of two entries contains five letters that are the same as that of the preceding entry, this can be signified by storing one character representing the number 5 and the remaining characters not common to both entries.

Abstract (fr)
Un procédé et un appareil permettent la compression de données alphanumériques qui sont stockées ou transmises sous forme de codes digitaux. Un dictionnaire est créé pour affecter chaque mot du texte alphanumérique et la ponctuation qui le suit à une adresse ou signe unique comprenant par exemple jusqu'à 16 bits (deux bytes). Chaque mot du texte alphanumérique est alors remplacé par l'adresse qui se réfère à ce mot du dictionnaire. Etant donné que le dictionnaire peut contenir jusqu'à $2^{16} = 65.536$ entrées, il est d'une taille plus que suffisante pour la mémorisation des mots associés à presque n'importe quel livre. Etant donné que seuls deux bytes d'informations sont nécessaires pour s'adresser à l'un quelconque de ces 65.000 mots, le remplacement de chaque mot de texte par deux bytes d'informations d'adresses réduit d'un facteur de trois environ le nombre moyen de chiffres requis pour stocker le texte. Des réductions supplémentaires de 25% ou davantage de la longueur du texte comprimé peuvent être obtenues dans la plupart des cas en représentant les mots utilisés le plus fréquemment par des signes qui sont plus courts que deux bytes en longueur. Le nombre de bytes requis pour mémoriser le dictionnaire peut être sensiblement réduit en mémorisant les mots dans l'ordre alphabétique et en tirant avantage de la redondance des caractères qui en résulte. Ainsi, si la seconde de deux entrées contient cinq lettres qui sont identiques à celles de l'entrée précédente, ceci peut être signifié en stockant un caractère représentant le nombre 5 et les autres caractères restants qui ne sont pas communs aux deux entrées.

IPC 1-7
G06F 9/36; **G06F 15/38**

IPC 8 full level
G06F 17/22 (2006.01); **H03M 7/30** (2006.01); **G06F 5/00** (2006.01); **H03M 7/42** (2006.01); **H04L 23/00** (2006.01)

CPC (source: EP)
G06F 40/12 (2020.01); **H03M 7/42** (2013.01)

Designated contracting state (EPC)
BE CH DE FR GB LI NL SE

DOCDB simple family (publication)
WO 8501814 A1 19850425; CA 1226369 A 19870901; EP 0160672 A1 19851113; EP 0160672 A4 19860512; IT 1180100 B 19870923; IT 8468039 A0 19841019; IT 8468039 A1 19860419; JP S61500345 A 19860227

DOCDB simple family (application)
US 8401667 W 19841017; CA 465602 A 19841017; EP 84903871 A 19841017; IT 6803984 A 19841019; JP 50381384 A 19841017