

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
28 April 2005 (28.04.2005)

PCT

(10) International Publication Number
WO 2005/039057 A1

(51) International Patent Classification⁷: **H03M 7/30**

(21) International Application Number:
PCT/AU2004/001406

(22) International Filing Date: 15 October 2004 (15.10.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2003905688 17 October 2003 (17.10.2003) AU

(71) Applicant (for all designated States except US):
PACBYTE SOFTWARE PTY LIMITED [AU/AU]; 36A
Langston Place, Epping, New South Wales 2121 (AU).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **PARKER, Bruce**,
[AU/AU]; 23 Beryl Avenue, Mt Colah, New South Wales
2079 (AU).

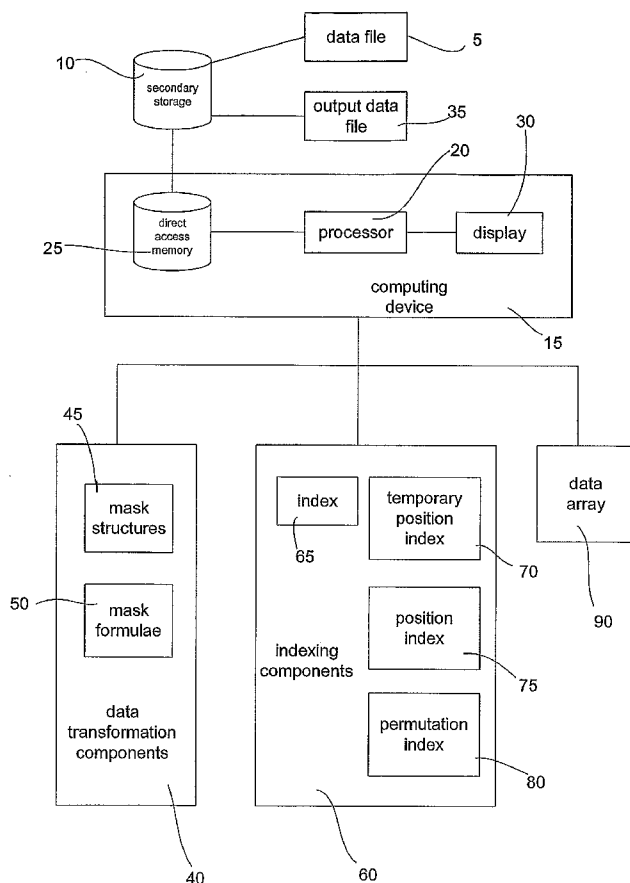
(74) Agents: **WEST-WALKER, Gregory, J** et al.; A J Park,
Level 11, 60 Marcus Clarke Street, Canberra, Australian
Capital Territory 2601 (AU).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: DATA COMPRESSION SYSTEM AND METHOD



(57) Abstract: The invention provides a method of compressing a data file comprising a sequence of bytes of a length greater than or equal to a predefined length, the method including the steps of: retrieving the data file from a secondary storage device; storing the data file in direct access memory; calculating the frequency of unique byte values within a sub-sequence of the data file, the sub-sequence having a length not exceeding the predefined length; creating an index for the sub-sequence, the index including a data value representing the calculated frequency of unique byte values within the sub-sequence; and on the sub-sequence having a frequency of unique byte values below a predetermined threshold, applying a data transformation to the sub-sequence to increase the frequency of unique byte values in the sub-sequence and adding to the index a data value representing the data transformation; on the sub-sequence having a frequency of unique byte values above a predefined threshold, adding to the index a data value representing the position of one or more unique values within the sub-sequence; creating an output data file, the data file having a file type identifier; and adding the index to the output data file.



FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *with international search report*