

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 June 2008 (05.06.2008)

PCT

(10) International Publication Number
WO 2008/067226 A1

(51) International Patent Classification:
G06F 7/24 (2006.01)

(74) Agent: **KOLODKA, Joseph**; 4 Independence Way, Suite 200, Princeton, New Jersey 08540 (US).

(21) International Application Number:
PCT/US2007/085357

(22) International Filing Date:
21 November 2007 (21.11.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/566,122 1 December 2006 (01.12.2006) US

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

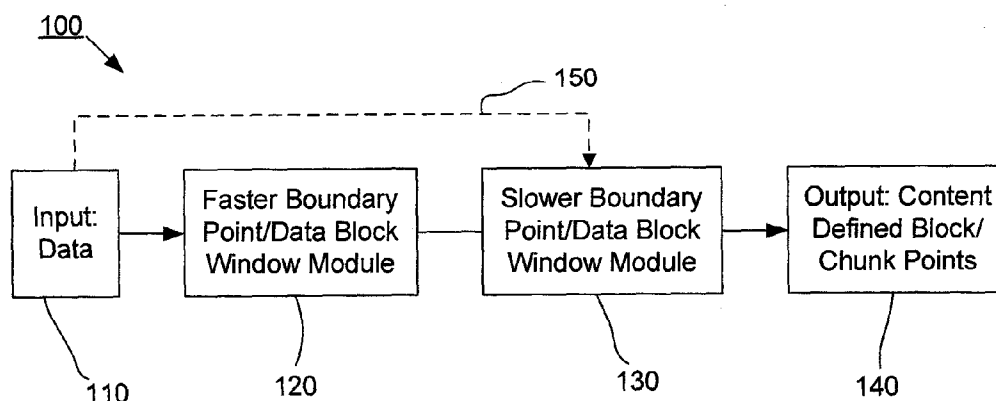
(71) Applicant (for all designated States except US): **NEC LABORATORIES AMERICA, INC.** [US/US]; 4 Independence Way, Suite 200, Princeton, New Jersey 08540 (US).

(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, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, 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).

(72) Inventors: **DUBNICKI, Cezary**; 1614 Mulberry Court, Monmouth Junction, New Jersey 08852 (US). **LICHOTA, Krzysztof**; A1. Ken 26/8, PL-02-797 Warszawa (PL). **KRUUS, Erik**; 261 Hamilton Road East, Hillsborough, New Jersey 08844 (US). **UNGUREANU, Cristian**; 147 Hamilton Avenue, Princeton, New Jersey 08540 (US).

Published:
— with international search report

(54) Title: METHODS AND SYSTEMS FOR DATA MANAGEMENT USING MULTIPLE SELECTION CRITERIA



(57) Abstract: Systems and methods for data management and data processing are provided. Embodiments may include systems and methods relating to fast data selection with reasonably high quality results, and may include a faster data selection function and a slower data selection function. Various embodiments may include systems and methods relating to data hashing and/or data redundancy identification and elimination for a data set or a string of data. Embodiments may include a first selection function is used to pre-select boundary points or data blocks/windows from a data set or data stream and a second selection function is used to refine the boundary points or data blocks/windows. The second selection function may be better at determining the best places for boundary points or data blocks/windows in the data set or data stream. In various embodiments, data may be processed by a first faster hash function and slower more discriminating second hash function.



WO 2008/067226 A1