



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2014/014186

(22) International Filing Date:
31 January 2014 (31.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/759,799 1 February 2013 (01.02.2013) US
13/827,558 14 March 2013 (14.03.2013) US

(71) Applicant: **AB INITIO TECHNOLOGY LLC** [US/US];
201 Spring Street, Lexington, Massachusetts 02421 (US).

(72) Inventors: **ISMAN, Marshall A.**; 11 Valleyspring Road,
Newton, Massachusetts 02158 (US). **EPSTEIN, Richard
Alan**; 863 Princes Point Road, Yarmouth, Maine 04096

(US). **HAUG, Ralf**; 340 Cleveland Ave, Largo, Florida
33770 (US). **ROBERTS, Andrew F.**; 78 Larchmont Road,
Melrose, Massachusetts 02176 (US). **RALSTON, John**;
542 Matilija Lane, Arroyo Grande, California 93420 (US).
RICHARDSON, John L.; 1421 Camino Meleno, Santa
Barbara, California 93111-1010 (US). **PNIOWER, Justin**;
39 Salutation Street, 4A, Boston, Massachusetts 02109
(US).

(74) Agents: **FEIGENBAUM, David L.** et al.; Fish & Richard-
son P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-
1022 (US).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,

[Continued on next page]

(54) Title: DATA RECORDS SELECTION

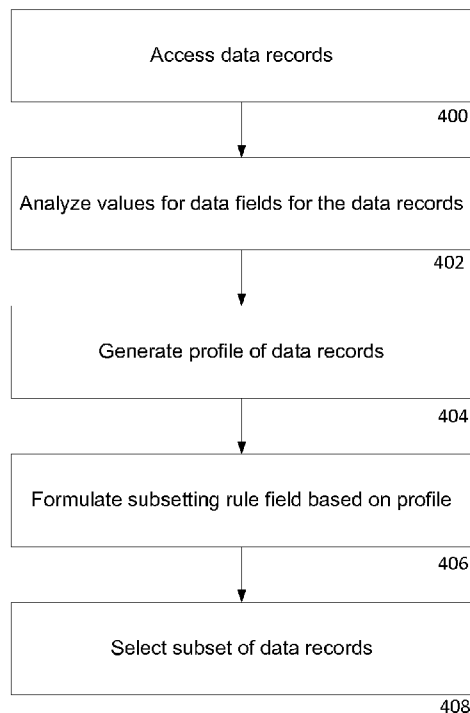


FIG. 4

(57) Abstract: A computer-implemented method includes accessing a plurality of data records, each data record having a plurality of data fields. The method further includes analyzing values for one or more of the data fields for at least some of the plurality of data records and generating a profile of the plurality of data records based on the analyzing. The method further includes formulating at least one subsetting rule based on the profile; and selecting a subset of data records from the plurality of data records based on the at least one subsetting rule.



MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

(88) Date of publication of the international search report:

9 October 2014