

(51) International Patent Classification:
G06N 5/02 (2006.01)(21) International Application Number:
PCT/US2015/048322(22) International Filing Date:
3 September 2015 (03.09.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/045,398 3 September 2014 (03.09.2014) US

(71) Applicant: THE DUN & BRADSTREET CORPORATION [US/US]; 103 Jfk Parkway, Short Hills, NJ 07078 (US).

(72) Inventors: SCRIFFIGNANO, Anthony, J.; 37 Woodrow Place, West Caldwell, NJ 07006 (US). SUNBHANICH, Yiern; 7372 Lavender Lane, Lewis Center, OH 43035 (US). DAVIES, Robin, Fry; 7649 N Territorial Road, Plymouth, MI 48170 (US). MATTHEWS, Warwick; 16 Bath Road, Glen Iris, Victoria, 3114 (AU).

(74) Agent: GREELEY, Paul, D.; Ohlandt, Greeley, Ruggiero & Perle, LLP, One Landmark Square, 10th Floor, Stamford, CT 06901 (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, LU, LY, MA, MD, ME, 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.

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

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: SYSTEM AND PROCESS FOR ANALYZING, QUALIFYING AND INGESTING SOURCES OF UNSTRUCTURED DATA VIA EMPIRICAL ATTRIBUTION

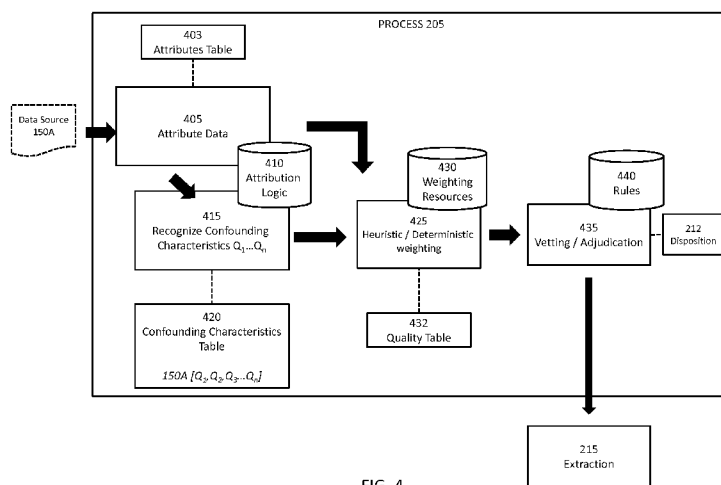


FIG. 4

(57) Abstract: There is provided a method that includes (a) receiving data from a data source, (b) attributing the data source in accordance with rules, thus yielding an attribute, (c) analyzing the data to identify a confounding characteristic in the data, (d) calculating a qualitative measure of the attribute, thus yielding a weighted attribute, (e) calculating a qualitative measure of the confounding characteristic, thus yielding a weighted confounding characteristic, (f) analyzing the weighted attribute and the weighted confounding characteristic, to produce a disposition, (g) filtering the data in accordance with the disposition, thus yielding extracted data, and (h) transmitting the extracted data to a downstream process. There is also provided a system that executes the method, and a storage device that contains instructions for controlling a processor to perform the method.