

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
21 October 2021 (21.10.2021)



(10) International Publication Number
WO 2021/209862 A1

- (51) **International Patent Classification:**
G06F 17/40 (2006.01)
- (21) **International Application Number:**
PCT/IB2021/052878
- (22) **International Filing Date:**
07 April 2021 (07.04.2021)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/011,990 17 April 2020 (17.04.2020) US
17/075,617 20 October 2020 (20.10.2020) US
- (71) **Applicant:** **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York, 10504 (US).
- (71) **Applicants (for MG only):** **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, North Harbour, Portsmouth Hampshire PO6 3AU (GB). **IBM (CHINA) INVESTMENT COMPANY LIMITED** [CN/CN]; 25/F, Pangu Plaza, No. 27, Central North 4th Ring Road, Chaoyang District, Beijing 100101 (CN).
- (72) **Inventors:** **SHRIVASTAVA, Shrey;** IBM Corporation, 1101 Kitchawan Road, PO Box 218, Yorktown Heights,

New York 10598-0218 (US). **BHAMIDIPATY, Anuradha;** IBM Corporation, 1101 Kitchawan Road, PO Box 218, Yorktown Heights, New York 10598-0218 (US). **PATEL, Dhavalkumar;** IBM Corporation, 1101 Kitchawan Road, PO Box 218, Yorktown Heights, New York 10598-0218 (US).

(74) **Agent:** **LITHERLAND, David;** IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Hursley, Winchester Hampshire SO21 2JN (GB).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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,

(54) **Title:** DYNAMIC DISCOVERY AND CORRECTION OF DATA QUALITY ISSUES

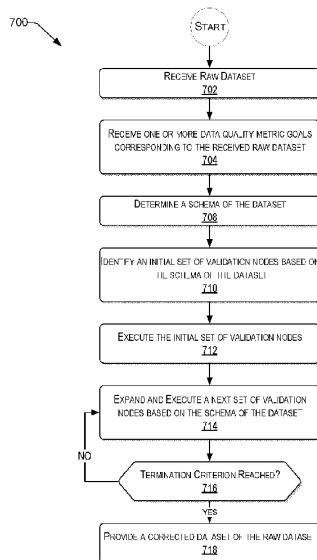


FIG. 7

(57) **Abstract:** A computing device, method, and system are provided of improving data quality to conserve computational resources. The computing device receives a raw dataset. One or more data quality metric goals corresponding to the received raw dataset are received. A schema of the dataset is determined. An initial set of validation nodes is identified based on the schema of the dataset. The initial set of validation nodes are executed. A next set of validation nodes are iteratively expanded and executed based on the schema of the dataset until a termination criterion is reached. A corrected dataset of the raw dataset is provided based on the iterative execution of the initial and next set of validation nodes.

WO 2021/209862 A1

[Continued on next page]

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).

Published:

- *with international search report (Art. 21(3))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*