

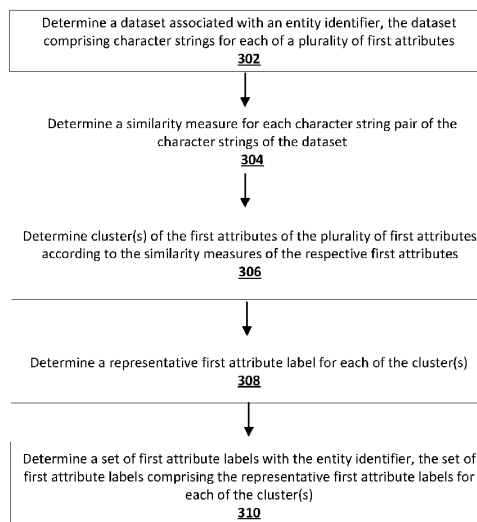


- (51) **International Patent Classification:**
G06F 16/903 (2019.01) *G06F 16/2457* (2019.01)
- (21) **International Application Number:**
PCT/NZ2024/050073
- (22) **International Filing Date:**
19 June 2024 (19.06.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
2023902046 28 June 2023 (28.06.2023) AU
- (71) **Applicant: XERO LIMITED** [NZ/NZ]; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ).
- (72) **Inventor: FEDYASHOV, Victor**; C/o- Xero Limited, 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ).
- (74) **Agent: FB RICE PTY LTD**; Level 33, 477 Collins St, Melbourne, Victoria 3000 (AU).
- (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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.

(54) **Title:** METHODS AND SYSTEMS FOR DETERMINING SETS OF SECOND ATTRIBUTES ASSOCIATED WITH RESPECTIVE FIRST ATTRIBUTES

300 ↘

Figure 3



(57) **Abstract:** Described embodiments relate to a method comprising determining a dataset associated with an entity attribute, the dataset comprising character strings for each of a plurality of first attributes. For each character string pair of the character strings of the dataset, the method comprises determining a similarity measure based on a lowest value of: (i) a minimum number of character edits required to convert the candidate character string into a substring of the target character string; and (ii) a minimum number of character edits required to convert the candidate character string into the target character string. The method further comprises determining cluster(s) of the first attributes according to their similarity measures, determining a representative label for each of the clusters, and determining a set of first attribute labels associated with the entity attribute comprising the representative labels for each of the cluster(s).

[Continued on next page]

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))*
- *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))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*