

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
28 September 2023 (28.09.2023)



(10) International Publication Number
WO 2023/182893 A1

(51) International Patent Classification:

G06F 40/284 (2020.01) G06N 20/00 (2019.01)
G06F 40/279 (2020.01) G06Q 20/40 (2012.01)

(21) International Application Number:

PCT/NZ2023/050011

(22) International Filing Date:

16 February 2023 (16.02.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2022900697 21 March 2022 (21.03.2022) AU

(71) Applicant: **XERO LIMITED** [NZ/NZ]; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ).

(72) Inventors: **DRIDAN, Rebecca**; C/- Xero Limited, 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ). **CHEAH, Soon-Ee**; C/- 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).

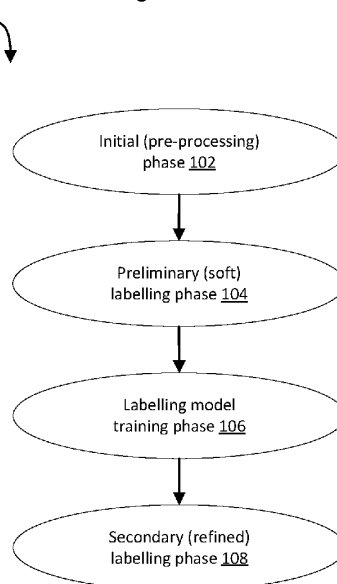
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, 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, 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, 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,

(54) Title: METHODS, SYSTEMS, AND COMPUTER-READABLE MEDIA FOR GENERATING LABELLED DATASETS

100

Figure 1



(57) Abstract: A method comprising determining token(s) from financial documents and determining a set of preliminary attribute labels for the token(s), wherein the set is associated with attribute type(s). The method further comprises providing the set for each token to an attribute prediction model to determine, for the token, a confidence value for each attribute type(s), determining subsets of token, each subset being associated with a respective document of the plurality of documents and determining a set of refined labels for each document based on the confidence values, wherein the set of refined labels comprises a value for attribute type(s).

[Continued on next page]



WO 2023/182893 A1

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))*
- *in black and white; the international application as filed
contained color or greyscale and is available for download
from PATENTSCOPE*