

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
25 August 2022 (25.08.2022)



(10) International Publication Number
WO 2022/177447 A1

- (51) International Patent Classification:
G06K 9/46 (2006.01)
- (21) International Application Number:
PCT/NZ2021/050133
- (22) International Filing Date:
19 August 2021 (19.08.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
2021900419 18 February 2021 (18.02.2021) AU
- (71) Applicant: **XERO LIMITED** [NZ/NZ]; 19-23 Taranaki Street, Te Aro, WELLINGTON, 6011 (NZ).
- (72) Inventors: **GLEYZES, Jerome**; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ). **KHODEIR, Mohamed**;

- 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ). **FAKHOURI, Salim**; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ). **WU, Yu**; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ). **CHEAH, Soon-Ee**; 19-23 Taranaki Street, Te Aro, Wellington, 6011 (NZ).
- (74) Agent: **FB RICE**; Level 33, 477 Collins Street, 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, 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,

(54) Title: SYSTEMS AND METHODS FOR GENERATING DOCUMENT NUMERICAL REPRESENTATIONS

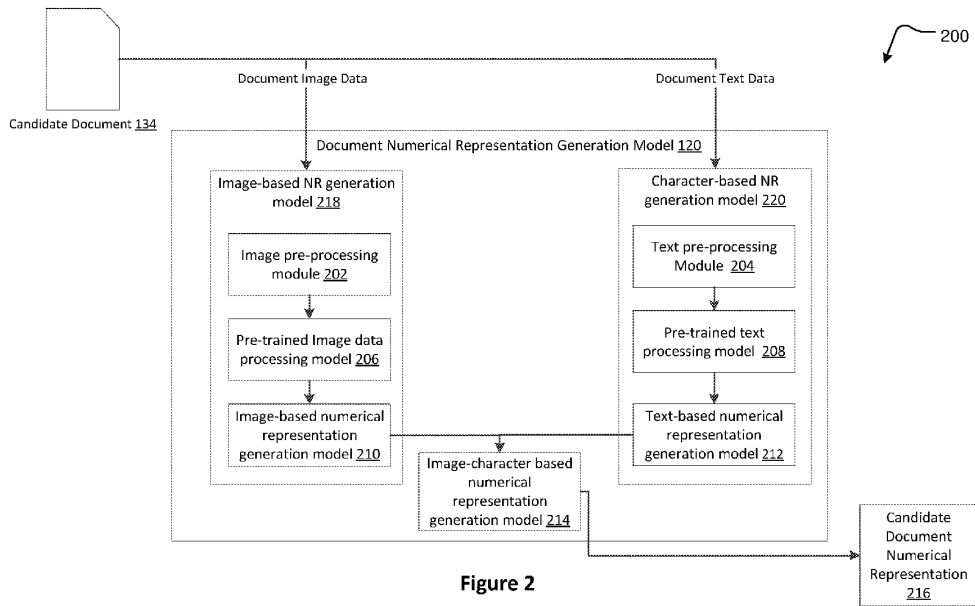


Figure 2

(57) Abstract: Described embodiments relate to a method comprising: determining a candidate document comprising image data and character data and extracting the image data and the character data from the candidate document. The method comprises providing, to an image-based numerical representation generation model, the image data, and generating, by the image-based numerical representation generation model, an image-based numerical representation of the image data. The method comprises providing, to a character-based numerical representation generation model, the character data; and generating, by the character-based numerical representation generation model, a character-based numerical representation of the character data. The method comprises providing, to a consolidated image-character based numerical representation generation model, the image-based numerical representation and the character-based numerical representation; and generating, by the consolidated image-character based numerical representation generation model, a

WO 2022/177447 A1

[Continued on next page]

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, 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*

combined image-character based numerical representation of the candidate document.