



- (51) **International Patent Classification:**
G06F 16/906 (2019.01) *G06F 16/93* (2019.01)
- (21) **International Application Number:**
PCT/US2024/040377
- (22) **International Filing Date:**
31 July 2024 (31.07.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
18/747,202 18 June 2024 (18.06.2024) US
- (63) **Related by continuation (CON) or continuation-in-part (CIP) to earlier application:**
US 18/747,202 (CON)
Filed on 18 June 2024 (18.06.2024)
- (71) **Applicant: VARONIS SYSTEMS, INC.** [US/US]; 1250 Broadway 28th Floor, New York, New York 10001 (US).
- (72) **Inventors: SNEH, Ron;** c/o Varonis Systems, Inc., 1250 Broadway 28th Floor, New York, New York 10001 (US).

NEYSTADT, John Eugene; c/o Varonis Systems, Inc., 1250 Broadway 28th Floor, New York, New York 10001 (US). **OSI, Amit;** c/o Varonis Systems, Inc., 1250 Broadway 28th Floor, New York, New York 10001 (US). **BASS, David;** c/o Varonis Systems, Inc., 1250 Broadway 28th Floor, New York, New York 10001 (US). **BELGI, Amir;** c/o Varonis Systems, Inc., 1250 Broadway 28th Floor, New York, New York 10001 (US).

(74) **Agent: GREEN, Edward H.;** Coats & Bennett, PLLC, 1400 Crescent Green, Suite 300, Cary, North Carolina 27518 (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, 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,

(54) **Title:** METHOD FOR AUTOMATICALLY CATEGORIZING DATA ITEMS

Creating initial classifications

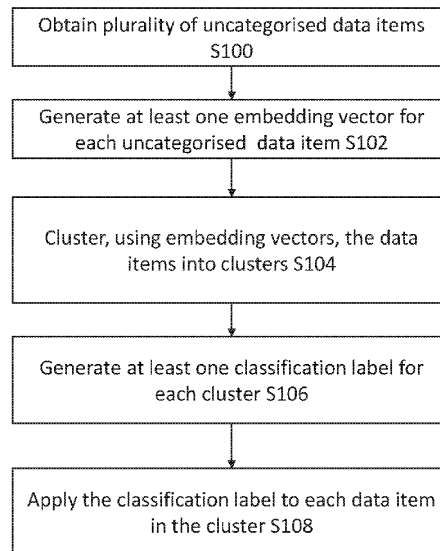


FIG. 3

(57) **Abstract:** Broadly speaking, the present techniques provide an automatic way of classifying data items within an environment (e.g. a business, workplace, organisation, etc.). This is advantageous over existing techniques which require manual classification of data items, which is time consuming in environments where hundreds of new data items may be generated in a day or week. The present techniques use an embedding machine learning, ML, model and an LLM to automatically determine the relevant classification label(s) for an unlabelled data item.

[Continued on next page]

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, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*