

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number
WO 2019/035765 A9

(43) International Publication Date
21 February 2019 (21.02.2019)

WIPO | PCT

(51) International Patent Classification:

G06N 20/00 (2019.01) G06F 17/28 (2006.01)
G06F 16/00 (2019.01) G06F 21/62 (2013.01)
G06F 17/27 (2006.01)

(72) Inventor: MUFFAT, Christopher; 298 Beach Road,
#27-14, Singapore 199554 (SG).

(74) Agent: SPRUSON & FERGUSON (ASIA) PTE LTD;
P.O. Box 1531, Robinson Road Post Office, Singapore
903031 (SG).

(21) International Application Number:

PCT/SG2018/050411

(22) International Filing Date:

14 August 2018 (14.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: DATHENA SCIENCE PTE. LTD. [SG/SG];
28C Stanley St, Singapore 068737 (SG).

(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, 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,

(54) Title: METHODS, MACHINE LEARNING ENGINES AND FILE MANAGEMENT PLATFORM SYSTEMS FOR CONTENT
AND CONTEXT AWARE DATA CLASSIFICATION AND SECURITY ANOMALY DETECTION

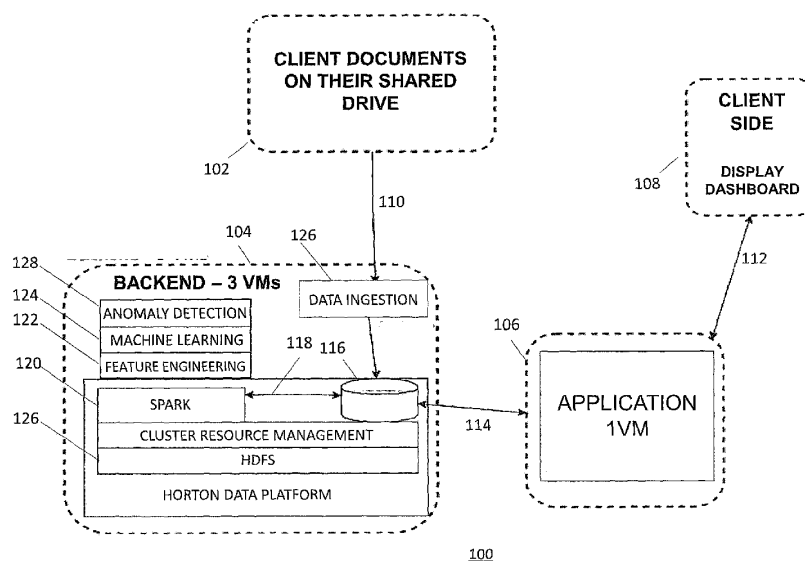


Fig. 1A

(57) Abstract: Systems, methods and computer readable medium are provided to perform a method for content and context aware data classification or a method for content and context aware data security anomaly detection. The method for content and context aware data confidentiality classification includes scanning one or more documents in one or more network data repositories of a computer network and extracting content features and context features of the one or more documents into one or more term frequency-inverse document frequency (TF-IDF) vectors and one or more latent semantic indexing (LSI) vectors. The method further includes classifying the one or more documents into a number of category classifications by machine learning the extracted content features and context features of the one or more documents at a file management platform of the computer network, each of the category classifications

[Continued on next page]

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

Declarations under Rule 4.17:

- *of inventorship (Rule 4.17(iv))*

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*

(48) Date of publication of this corrected version:

21 March 2019 (21.03.2019)

(15) Information about Correction:

see Notice of 21 March 2019 (21.03.2019)