

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
12 November 2020 (12.11.2020)



(10) International Publication Number
WO 2020/227211 A1

(51) International Patent Classification:

G06F 16/21 (2019.01) G06N 20/10 (2019.01)
G06F 16/27 (2019.01) G06N 20/20 (2019.01)

(21) International Application Number:

PCT/US2020/031319

(22) International Filing Date:

04 May 2020 (04.05.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/842,569 03 May 2019 (03.05.2019) US

(71) Applicant: **VISA INTERNATIONAL SERVICE ASSOCIATION** [US/US]; P.O. Box 8999, San Francisco, California 94128 (US).

(72) Inventors: **KOLLAMPARAMBIL, Tony, Thomas**; 39469 Gallaudet Dr., Apt. 109, Fremont, California 94538 (US). **NOSSEIR, Mohamed**; 4610 Mira Loma Street, Castro Valley, California 94546 (US).

(74) Agent: **PREPELKA, Nathan, J.** et al.; The Webb Law Firm, 420 Ft. Duquesne Blvd., Suite 1200, Pittsburgh, Pennsylvania 15222 (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, 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: SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR GENERATING ENHANCED N-GRAM MODELS

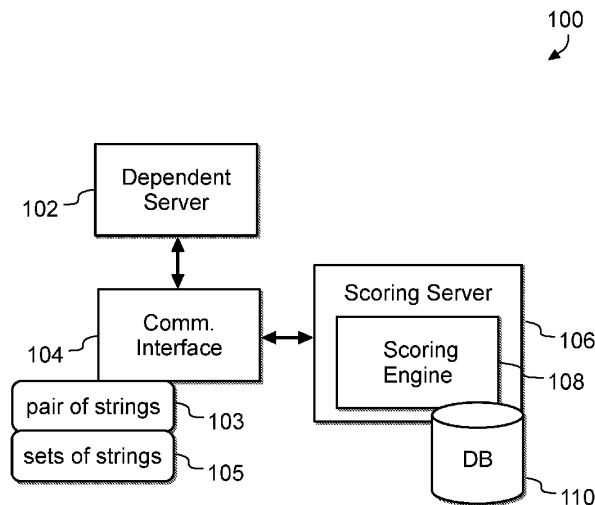


FIG. 1

(57) Abstract: A method, system, and computer program product is provided for generating enhanced n-gram models for use with monitoring systems. The method includes determining that a leading pair of characters of a first data string does not match a leading pair of characters of a second data string and inserting a placeholder character at a first-index position in each data string. The method further includes inserting a placeholder character between each character pair of the first data string in which a first character matches a character of the second data string at a same index position and in which a second character matches a character of the second data string at an index position immediately following a same index position, and generating a similarity score based on the length of the data strings and triggering a remedial process in response to the similarity score exceeding a predetermined threshold.

[Continued on next page]

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).

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))*