

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
21 September 2023 (21.09.2023)



(10) International Publication Number
WO 2023/177401 A1

(51) International Patent Classification:
G06F 17/27 (2006.01)

(21) International Application Number:
PCT/US2022/020792

(22) International Filing Date:
17 March 2022 (17.03.2022)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **SMARSH, INC.** [US/US]; 851 SW 6th Ave,
Ste 800, Portland, Oregon 97204 (US).

(72) Inventors: **NIEBOROWSKI, Thomas**; 671 Wagon Rd.,
Sebastopol, California 95472 (US). **GOND, Shravan Ku-**
mar; SA 20/4, K-21, Ruppenpur, Natui, Sarnath, Varanasi,
Uttar Pradesh 221007 (IN). **NEELI, Ramakrishna**; 001
RKR Vilas, 4th Main, 6th Cross, Konena Agrahar, Old

Airport Rd, Bangalore, Karnataka 56001 (IN). **HENRY,**
George; 2128 SE 35th Pl, Portland, Oregon 97214 (US).

(74) Agent: **THOMPSON, Adam J.**; 1600 Atlanta Financial
Center, 3343 Peachtree Road, NE, Atlanta, Georgia (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, IT, JM, 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, 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.

(54) Title: SYSTEMS AND PROCESSES FOR NATURAL LANGUAGE PROCESSING

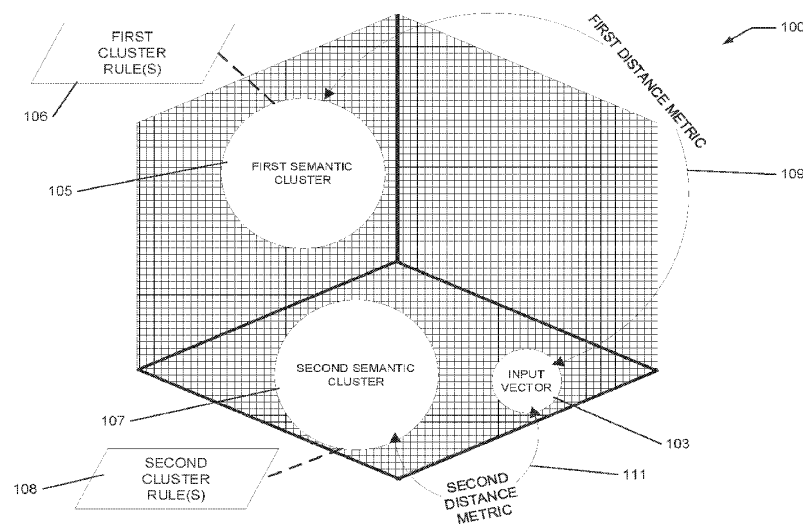


FIG. 1

(57) Abstract: A system for natural language processing includes a memory and at least one computing device in communication with the memory. The at least one computing device can receive a plurality of first data items and generate a cluster based on the plurality of first data items. The at least one computing device can intercept a plurality of second data items communicated between a first computing device and at least one second computing device. The at least one computing device can generate at least one vector based on the plurality of second data items and determine a similarity score between the at least one vector and the cluster. The at least one computing device can identify at least one of the plurality of second data items for review based at least in part on the similarity score.

[Continued on next page]

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, 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, 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))