



- (51) **International Patent Classification:**
G06F 16/31 (2019.01) *G06F 16/35* (2019.01)
- (21) **International Application Number:**
PCT/US2024/012376
- (22) **International Filing Date:**
22 January 2024 (22.01.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/440,966 25 January 2023 (25.01.2023) US
- (71) **Applicants:** **DOW GLOBAL TECHNOLOGIES LLC** [US/US]; 2211 H.H. Dow Way, Midland, Michigan 48674 (US). **ROHM AND HAAS COMPANY** [US/US]; 100 North Independence Mall West, Philadelphia, Pennsylvania 19106 (US).
- (72) **Inventors:** **WANG, Zhenyu**; 332 SH 332 E, APB-4A, 134C, Lake Jackson, Texas 77566 (US). **STEVENSON, Dale**; 2 Manse Road, West Hill, Scarborough, Ontario M1E 3T9 (CA). **CHIANG, Leo H.**; 230 Abner Jackson Pkwy., ECB 1A114, Lake Jackson, Texas 77566 (US). **CASTIL-**

LO, Ivan; 332 SH 332 E, APB-4A, 134B, Lake Jackson, Texas 77566 (US). **SCHMIDT, Alix**; 2211 H.H. Dow Way, Midland, Michigan 48674 (US). **BRAUN, Birgit**; 332 SH 332 E, Lake Jackson, Texas 77566 (US).

(74) **Agent:** **LORFING, Jason P.**; Brooks, Cameron & Huebsch, PLLC, 1201 Marquette Avenue S, Suite 400, Minneapolis, Minnesota 55403 (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, 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,

(54) **Title:** INFORMATION EXTRACTION FROM DOMAIN-SPECIFIC DOCUMENTS

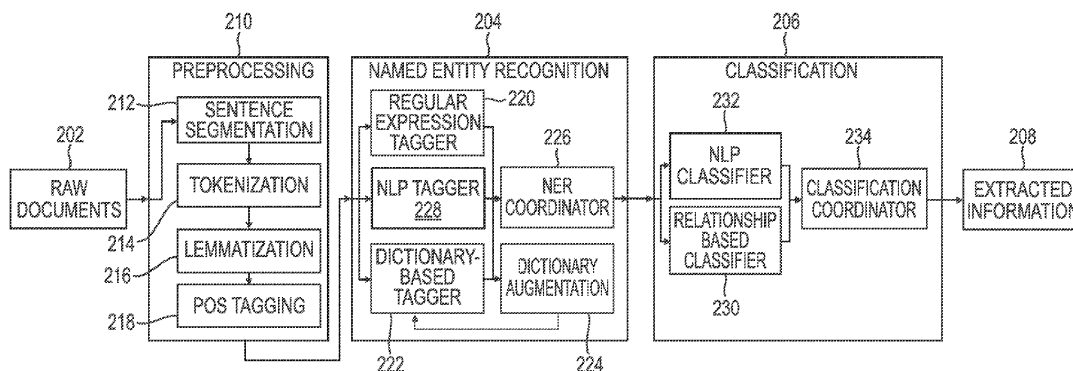


Fig. 2

(57) **Abstract:** Information extraction from domain-specific documents can include performing named entity recognition thereon. A regular expression tagger can tag regular expressions in the domain-specific documents without pre-annotated training sets for the regular expression tagger. Defined domain-specific terms can be augmented with terms and phrases related to the defined domain-specific terms based on rules specified for desired information about named entities to be extracted. A dictionary-based tagger can tag the defined and augmented domain-specific terms and phrases in the plurality of domain-specific documents. Conflicts of named entity recognition between the regular expression tagger and the dictionary-based tagger can be resolved. The desired information about named entities can be extracted. Classification can be performed to identify whether the domain-specific documents are related to the desired information.

[Continued on next page]

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).

Published:

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