



- (51) International Patent Classification:
G06F 17/00 (2019.01)
- (21) International Application Number:
PCT/US2020/065361
- (22) International Filing Date:
16 December 2020 (16.12.2020)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/948,696 16 December 2019 (16.12.2019) US
- (71) Applicant: **TRIALMATCH.ME, INC. D/B/A/TRIAL-JECTORY** [US/US]; 700 Canal Street, Stamford, CT 06902 (US).
- (72) Inventors: **BADER, Tzvia**; 86 Maple Ave, Closter, NJ 07624 (US). **GILDOR, Guy**; 10 Oriel Ofek St., App 4, Herzliya (IL).
- (74) Agent: **HUESTIS, Erik, A.** et al.; Foley Hoag LLP, 155 Seaport Boulevard, Boston, MA 02210-2600 (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, 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: UNSUPERVISED TAXONOMY EXTRACTION FROM MEDICAL CLINICAL TRIALS

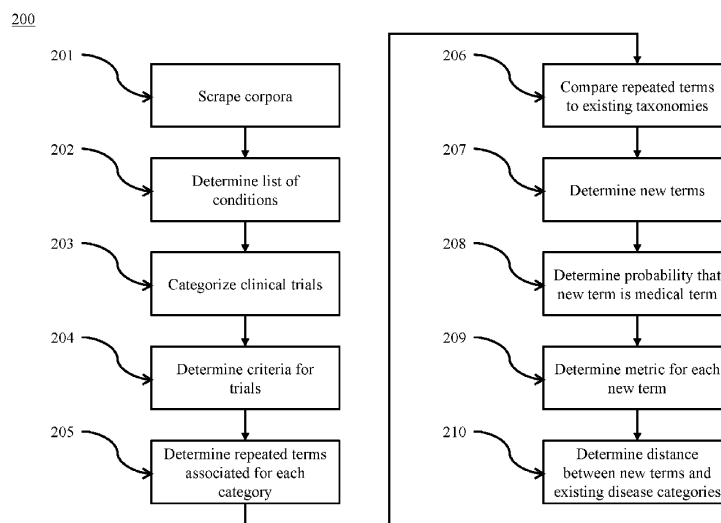


Fig. 2

(57) **Abstract:** Unsupervised taxonomy extraction from medical clinical trials for supplementing a keyword mapping to disease conditions is provided. One or more corpus of clinical trial descriptions is read. A list of disease conditions is read. A plurality of categories of clinical trials is determined. A frequency of occurrence for each of one or more repeated terms from each category of clinical trials is determined. A set of category-specific repeated terms is determined. A set of new terms is determined. A plurality of vectors for each new term in the set of new terms is determined where each vector for a particular new term corresponding to the one or more associated keyword. One or more of the new terms in the set of new terms is selected based on the vectors. Each of the selected new terms is mapped to a disease condition thereby generating a supplemented list of disease conditions.

[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, 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))