

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
23 December 2021 (23.12.2021)

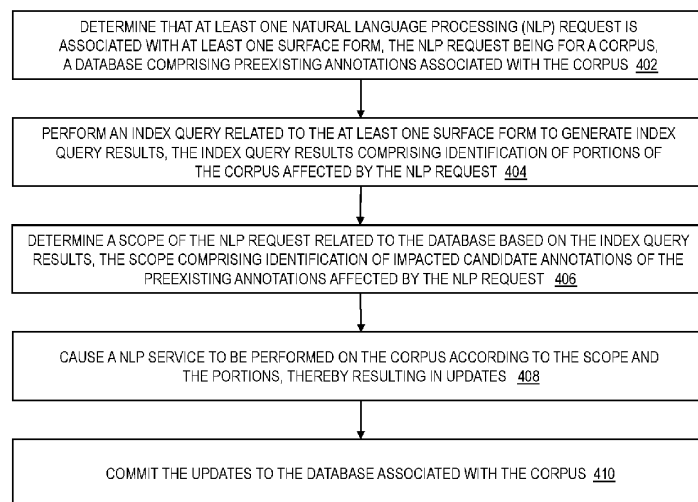


(10) International Publication Number
WO 2021/254227 A1

- (51) International Patent Classification:
G06F 16/332 (2019.01)
- (21) International Application Number:
PCT/CN2021/099122
- (22) International Filing Date:
09 June 2021 (09.06.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
16/904,924 18 June 2020 (18.06.2020) US
- (71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (71) Applicant (for MG only): **IBM (CHINA) CO., LIMITED** [CN/CN]; 7F, Bldg. 10, ZhangJiang Innovation Park, 399 Keyuan Road, ZhangJiang High-Tech Campus, Pudong New Area, Shanghai 201203 (CN).
- (72) Inventors: **CARRIER, Scott**; 4205 S Miami Blvd., Durham, NC 27703 (US). **BULL, Brendan**; 4205 S Miami Blvd., Durham, NC 27703-9141 (US). **FELT, Paul, Lewis**; 2889 Ashton Blvd., Lehi, UT 84043 (US). **MANSJUR, Dwi Sianto**; c/o 650 Harry Road Almaden Research Center, San Jose, CA 95120-6099 (US).
- (74) Agent: **LIU, SHEN & ASSOCIATES**; 10th Floor, Building 1, 10 Caihefang Road, Haidian District, Beijing 100080 (CN).
- (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,

(54) Title: TARGETED PARTIAL RE-ENRICHMENT OF A CORPUS BASED ON NLP MODEL ENHANCEMENTS

FIG. 4



(57) Abstract: A computer-implemented method include determining that at least one natural language processing (NLP) request is associated with at least one surface form, the NLP request being for a corpus, a database comprising preexisting annotations associated with the corpus. An index query related to the at least one surface form is performed to generate index query results, the index query results including identification of portions of the corpus affected by the NLP request. A scope of the NLP request related to the database is determined based on the index query results, the scope including identification of impacted candidate annotations of the preexisting annotations affected by the NLP request. An NLP service is performed on the corpus according to the scope and the portions, thereby resulting in updates. The updates are committed to the database associated with the corpus.

[Continued on next page]

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.

(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))