

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
05 March 2020 (05.03.2020)



(10) International Publication Number
WO 2020/047317 A1

- (51) **International Patent Classification:**
G06F 16/22 (2019.01) *G06F 16/31* (2019.01)
- (21) **International Application Number:**
PCT/US2019/048909
- (22) **International Filing Date:**
29 August 2019 (29.08.2019)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
201811014383.2 31 August 2018 (31.08.2018) CN
16/554,122 28 August 2019 (28.08.2019) US
- (71) **Applicant: ALIBABA GROUP HOLDING LIMITED**
[—/US]; Fourth Floor, One Capital Place, P.O. Box 847,
George Town, Grand Cayman (KY).
- (72) **Inventor: JIANG, Zhiyong;** Alibaba Group Legal Department, 5/F, Building 3, No. 969 West Wen Yi Road, Yu Hang District, Hangzhou, Zhejiang 311121 (CN).
- (74) **Agent: YAO, Shun;** 2800 Fifth Street, Suite 110, Davis, California 95618 (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, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) **Title:** SYSTEM AND METHOD FOR FACILITATING EFFICIENT INDEXING IN A DATABASE SYSTEM

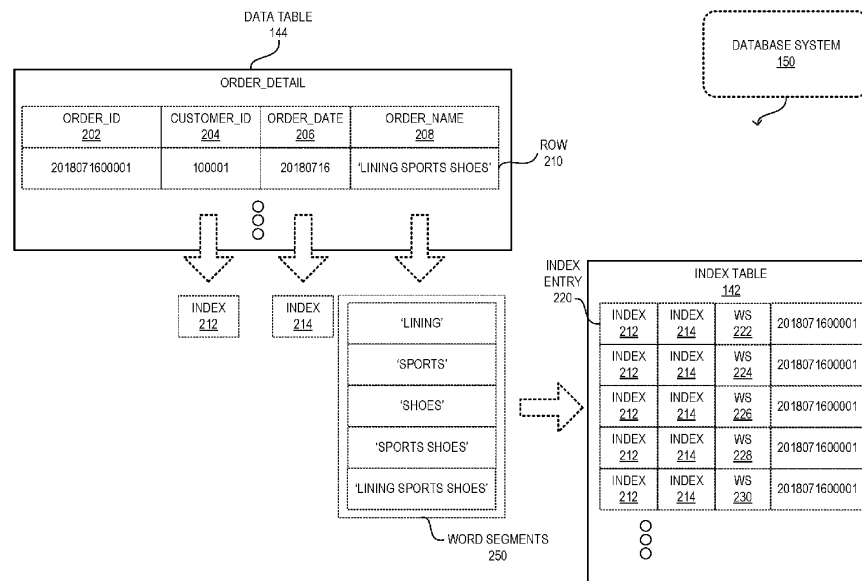


FIG. 2

(57) **Abstract:** Embodiments described herein provide a system for facilitating efficient indexing in a database. The system can maintain a data table in the database. The data table includes a first column for structured data and a second column for unstructured text. A row of the data table can include a first piece of data in the first column and a second piece of data in the second column. During operation, the system performs word segmentation on the second piece of data to generate a first set of semantically meaningful word segments. The system then generates a set of index entries that corresponds to the row. A respective index entry includes an index generated from the first piece of data and a corresponding word segment of the first set of semantically meaningful word segments.

[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))