



(51) International Patent Classification:
G06K 9/00 (2006.01)

(21) International Application Number:
PCT/IB2021/050648

(22) International Filing Date:
28 January 2021 (28.01.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/811,421 06 March 2020 (06.03.2020) US

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).

(71) Applicants (for MG only): **IBM ISRAEL - SCIENCE & TECHNOLOGY LTD.** [IL/IL]; Haifa University Campus, Mount Carmel, 31905 Haifa (IL). **IBM (CHINA) INVESTMENT COMPANY LTD.** [CN/CN]; 25/F, Pangu Plaza, No. 27, Central North 4th Ring Road, Chaoyang District, Beijing 100101 (CN).

(72) Inventors: **STAAR, Peter**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **DOLFI, Michelle**; IBM Research GmbH, Saumerstrasse 4,

Rueschlikon, 8803 Zurich (CH). **AUER, Christoph**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **GEORGOPOULOS, Leonidas**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **KAESTNER, Ralf**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **VELIZHEV, Alexander**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **NOGUER HIDALGO, Dal**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **KUZNETSOVA, Rita**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH). **BEKAS, Konstantinos**; IBM Research GmbH, Saumerstrasse 4, Rueschlikon, 8803 Zurich (CH).

(74) Agent: **GILBOA, Eyal**; IBM Israel - Science & Technology Ltd., Haifa University Campus, Mount Carmel, 31905 Haifa (IL).

(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,

(54) Title: EFFICIENT GROUND TRUTH ANNOTATION

(57) Abstract: A computer-implemented method for determining a set of target items to be annotated for training a machine learning application. The method comprises providing a training data set with a set of data samples and an auto-encoder with a classifier. The auto-encoder comprises an embedding model that maps the set of data samples to a set of compressed feature vectors. The set of compressed feature vectors define a compressed feature matrix. Further provided are: a definition of a graph associated to the compressed feature matrix, applying a clustering-algorithm to identify node clusters of the graph and applying a centrality algorithm to identify central nodes of the node clusters, retrieving from an annotator node labels for the central nodes, propagating the annotated node labels to other nodes of the graph and performing a training of the embedding model and the classifier with the annotated and the propagated node labels.

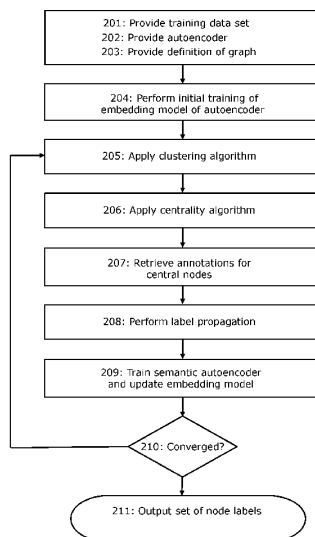


FIG. 2a

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))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*