



(51) International Patent Classification:
G06F 40/295 (2020.01)

(21) International Application Number:
PCT/CN2020/086780

(22) International Filing Date:
24 April 2020 (24.04.2020)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.** [CN/CN]; No. 556 Xixi Road, 8th Floor, Section B, Suite 801-11, West Lake District, Hangzhou, Zhejiang 310000 (CN).

(72) Inventors: **WANG, Desheng**; No. 556 Xixi Road, 8th Floor, Section B, Suite 801-11, West Lake District, Hangzhou, Zhejiang 310000 (CN). **LIU, Jiawei**; No. 556 Xixi Road, 8th Floor, Section B, Suite 801-11, West Lake District, Hangzhou, Zhejiang 310000 (CN). **ZHANG, Peng**; No. 556 Xixi Road, 8th Floor, Section B, Suite 801-11, West Lake District, Hangzhou, Zhejiang 310000 (CN).

(74) Agent: **BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**; Room 409, Tower B, Ka

Wah Building, No. 9 Shangdi 3rd Street, Haidian District, Beijing 100085 (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, 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, 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).

(54) Title: NAMED ENTITY RECOGNITION AND EXTRACTION USING GENETIC PROGRAMMING

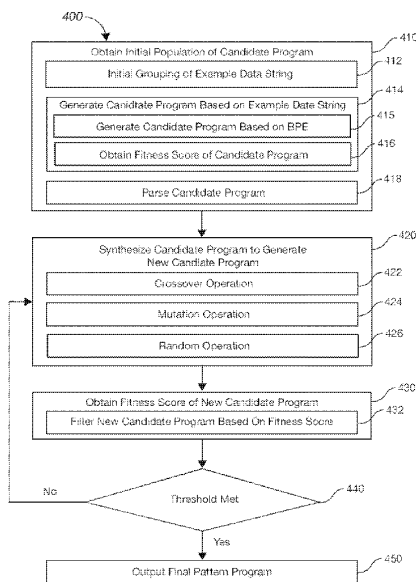


FIG. 4

(57) Abstract: Disclosed herein are methods, systems, and apparatus, including computer programs encoded on computer storage media, for generating a pattern program using a genetic algorithm. The genetic algorithm operates on example data strings that represent the data categories to be recognized or extracted through named entity recognition. In the initialization stage, the initial pattern programs are generated based on example data strings that represent the data categories to be recognized or extracted through named entity recognition. Starting from the initial pattern programs, genetic operations are iteratively conducted to generate generations of offspring pattern programs. In each round of the genetic operation, offspring pattern programs are generated through the crossover operation and the mutation operation.

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a)*

(88) Date of publication of the international search report:

25 February 2021 (25.02.2021)