

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2020/211250 A1

(43) 国际公布日
2020 年 10 月 22 日 (22.10.2020)

(51) 国际专利分类号:
G06F 17/27 (2006.01) G16H 10/60 (2018.01)

(21) 国际申请号: PCT/CN2019/103379

(22) 国际申请日: 2019 年 8 月 29 日 (29.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910316061.1 2019 年 4 月 19 日 (19.04.2019) CN

(72) 发明人: 丁佳佳 (DING, Jiajia); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

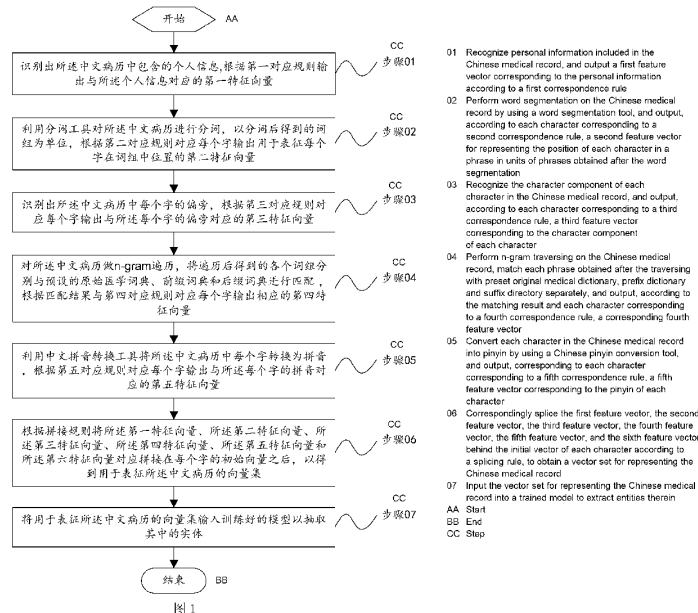
(74) 代理人: 北京英特普罗知识产权代理有限公司 (INTELLECPRO CHINA LIMITED); 中国北京市西城区车公庄大街 9 号五栋大楼 C 座 11 层, Beijing 100044 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): 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,

(71) 申请人: 平安科技 (深圳) 有限公司 (PING AN TECHNOLOGY(SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(54) Title: ENTITY RECOGNITION METHOD AND APPARATUS FOR CHINESE MEDICAL RECORD, DEVICE AND STORAGE MEDIUM

(54) 发明名称: 中文病历的实体识别方法、装置、设备及存储介质



(57) Abstract: An entity recognition method for a Chinese medical record, relating to the field of natural language processing. The method comprises the following steps: outputting, according to different correspondence rules, different feature vectors, comprising a first feature vector corresponding to personal information, a second feature vector for representing the position of each character in a phrase, a third feature vector corresponding to the character component of each character, a corresponding output fourth feature vector corresponding to each character, and a fifth feature vector corresponding to the pinyin of each character; then, correspondingly

[见续页]

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。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (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)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

splicing the feature vectors behind the initial vector of each character according to a splicing rule to obtain a vector set for representing a Chinese medical record; and finally, inputting the vector set for representing the Chinese medical record into a trained model to extract entities therein.

(57) 摘要: 一种中文病历的实体识别方法, 属于自然语言处理领域。该方法包括以下步骤: 根本不同的对应规则输出各种特征向量, 包括与所述个人信息对应的第一特征向量、用于表征每个字在词组中位置的第二个特征向量、与每个字的偏旁对应的第三个特征向量、对应每个字输出相应的第四个特征向量以及与所述每个字的拼音对应的第五个特征向量; 再根据拼接规则将各个特征向量对应拼接在每个字的初始向量之后, 以得到用于表征所述中文病历的向量集; 最后将用于表征所述中文病历的向量集输入训练好的模型以抽取其中的实体。