

(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 2023/226766 A1

(43) 国际公布日
2023 年 11 月 30 日 (30.11.2023)

(51) 国际专利分类号:
G06F 16/36 (2019.01)

(21) 国际申请号: PCT/CN2023/093275

(22) 国际申请日: 2023 年 5 月 10 日 (10.05.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210582554.1 2022 年 5 月 26 日 (26.05.2022) CN

(71) 申请人: 支付宝 (杭州) 信息技术有限公司 (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(72) 发明人: 祝慧佳 (ZHU, Huijia); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: DATA PROCESSING METHOD, APPARATUS AND DEVICE

(54) 发明名称: 数据处理方法、装置及设备

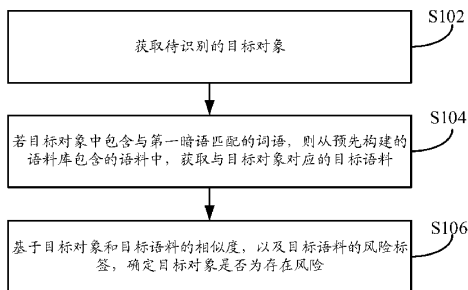


图 1A

S102 Acquire a target object to be identified
S104 If the target object includes a word matching a code word, acquire, from linguistic data included in a pre-constructed corpus, target linguistic data corresponding to the target object
S106 On the basis of the similarity between the target object and the target linguistic data, as well as a risk label of the target linguistic data, determine whether the target object is at risk

(57) Abstract: Provided in the embodiments of the present description are a data processing method, apparatus and device. The method comprises: acquiring a target object to be identified; if the target object includes a word matching a code word, acquiring, from linguistic data included in a pre-constructed corpus, target linguistic data corresponding to the target object, wherein the pre-constructed corpus comprises first linguistic data, the first linguistic data is risk linguistic data which is constructed on the basis of a second code word and target risk linguistic data, and the target risk linguistic data includes risk words with a preset association relationship with the second code word; and on the basis of the similarity between the target object and the target linguistic data, as well as a risk label of the target linguistic data, determining whether the target object is at risk.

(57) 摘要: 本说明书实施例提供了一种数据处理方法、装置及设备, 其中, 该方法包括: 获取待识别的目标对象; 若所述目标对象中包含与第一暗语匹配的词语, 则从预先构建的语料库包含的语料中, 获取与目标对象对应的目标语料, 所述预先构建的语料库包括第一语料, 所述第一语料为基于第二暗语和目标风险语料构建的风险语料, 所述目标风险语料包含与第二暗语具有预设关联关系的风险词语; 基于所述目标对象和所述目标语料的相似度, 以及所述目标语料的风险标签, 确定所述目标对象是否为存在风险。

[见续页]

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) 指定国(除另有指明，要求每一种可提供的地区
保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, 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, ME, 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))。