

(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 2019/223259 A1

(43) 国际公布日
2019 年 11 月 28 日 (28.11.2019) WIPO | PCT

- (51) 国际专利分类号:
G06F 17/27 (2006.01)
- (21) 国际申请号: PCT/CN2018/116037
- (22) 国际申请日: 2018 年 11 月 16 日 (16.11.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810515382.X 2018 年 5 月 25 日 (25.05.2018) CN
- (71) 申请人: 腾讯音乐娱乐科技(深圳)有限公司 (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。
- (72) 发明人: 冯穗豫 (FENG, Suiyu); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。
- (74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (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,

(54) Title: METHOD AND APPARATUS FOR PROCESSING WORD BANK

(54) 发明名称: 一种对词库进行处理的方法及装置

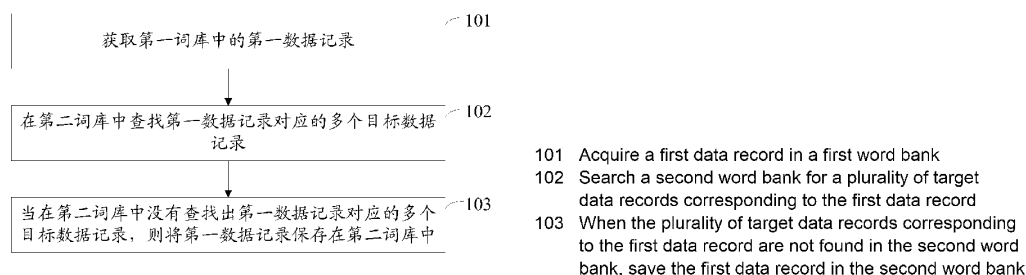


图 1

(57) Abstract: A method and apparatus for processing a word bank, which fall within the field of computers. The method comprises: acquiring a first data record in a first word bank (101), wherein the first data record comprises a multi-character entry and a first kana set corresponding to each Chinese character in the multi-character entry, and the first kana set corresponding to the Chinese character comprises at least one kana corresponding to the Chinese character; searching a second word bank for a plurality of target data records corresponding to the first data record (102), wherein target entries in each target data record are different constituent parts of the multi-character entry, the target entries in each target data record form the multi-character entry, and a second kana set corresponding to each Chinese character in the target entries in the target data record is respectively the same as the first kana set corresponding to each Chinese character; and when the plurality of target data records corresponding to the first data record are not found in the second word bank, saving the first data record in the second word bank (103). The method can improve kana labelling efficiency.

(57) 摘要: 一种对词库进行处理的方法及装置, 属于计算机领域。所述方法包括: 获取第一词库中的第一数据记录 (101), 第一数据记录包括一个多词条和该多词条中的每个汉字对应的第一假名集合, 汉字对应的第一假名集合包括该汉字对应的至少一个假名; 在第二词库中查找第一数据记录对应的多个目标数据记录 (102), 每个目标数据记录中的目标词条是该多词条的不同组成部分, 每个目标数据记录中的目标词条组成该多词条, 目标数据记录中的目标词条中的每个汉字对应的第二假名集合分别与每个汉字对应的第一假名集合相同; 当在第二词库中没有查找出第一数据记录对应的多个目标数据记录, 则将第一数据记录保存在所述第二词库中 (103)。该方法能够提高标注假名的效率。

[见续页]

WO 2019/223259 A1

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。

- (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))。