

(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 2022/151915 A1

(43) 国际公布日
2022 年 7 月 21 日 (21.07.2022)

- (51) 国际专利分类号:
G06F 40/284 (2020.01) *G06N 3/04* (2006.01)
- (21) 国际申请号: PCT/CN2021/139098
- (22) 国际申请日: 2021 年 12 月 17 日 (17.12.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110053835.3 2021 年 1 月 15 日 (15.01.2021) CN
- (71) 申请人: 北京沃东天骏信息技术有限公司 (BEIJING WODONG TIANJUN INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。北京京东世纪贸易有限公司 (BEIJING JINGDONG CENTURY TRADING CO., LTD.) [CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 201 室, Beijing 100176 (CN)。

- (72) 发明人: 李浩然 (LI, Haoran); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。袁鹏 (YUAN, Peng); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。
- (74) 代理人: 北京唯智勤实知识产权代理有限公司 (普通合伙) (BEIJING WISDOM INTELLECTUAL PROPERTY LAW FIRM); 中国北京市丰台区郭公庄中街 20 号北京方向 B 座 602 号, Beijing 100070 (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, 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, SA, SC, SD, SE, SG, SK, SL,

(54) Title: TEXT GENERATION METHOD AND APPARATUS, AND ELECTRONIC DEVICE AND COMPUTER-READABLE MEDIUM

(54) 发明名称: 文本生成方法、装置、电子设备和计算机可读介质

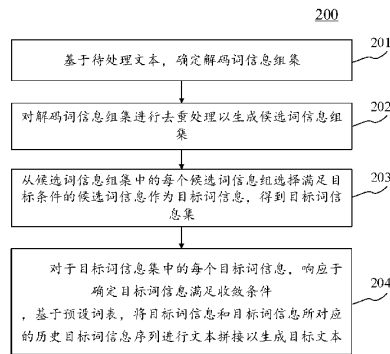


图 2

- 201 Determine a decoded-word information group set on the basis of text to be processed
- 202 De-duplicate the decoded-word information group set, so as to generate a candidate-word information group set
- 203 Select from each candidate-word information group in the candidate-word information group set, candidate-word information that meets a target condition, and take same as target-word information, so as to obtain a target-word information set
- 204 For each piece of target-word information in the target-word information set, in response to determining that the target-word information meets a convergence condition, perform, on the basis of a preset word table, text combining on the target-word information and a historical target-word information sequence corresponding to the target-word information, so as to generate target text

(57) Abstract: Disclosed are a text generation method and apparatus, and an electronic device and a computer-readable medium. A particular embodiment of the method comprises: determining a decoded-word information group set on the basis of text to be processed, wherein the text to be processed is used for describing a specified object; de-duplicating the decoded-word information group set, so as to generate a candidate-word information group set; selecting, from each candidate-word information group in the candidate-word information group set, candidate-word information that meets a target condition, and taking same as target-word information, so as to obtain a target-word information set; for each piece of target-word information in the target-word information set, in response to determining that the target-word information meets a convergence condition, combining, on the basis of a preset word table, the target-word information and a historical target-word information sequence corresponding to the target-word information, so as to generate target text. By means of the embodiment, the efficiency and accuracy of generating a text topic are improved.

[见续页]

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, 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))。

(57) 摘要: 本公开的实施例公开了文本生成方法、装置、电子设备和计算机可读介质。该方法的一具体实施方式包括: 基于待处理文本, 确定解码词信息组集, 其中, 待处理文本用于描述指定物体; 对解码词信息组集进行去重处理以生成候选词信息组集; 从候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息, 得到目标词信息集; 对于目标词信息集中的每个目标词信息, 响应于确定目标词信息满足收敛条件, 基于预设词表, 将目标词信息和目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。该实施方式提高了生成文本主题的效率及准确度。