(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

(43) 国际公布日
2024 年 8 月 22 日 (22.08.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/169529 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01)

(21) 国际申请号: PCT/CN2024/073350

(22) 国际申请日: 2024 年 1 月 19 日 (19.01.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310166877.7 2023年2月13日 (13.02.2023) CN

(71) 申请人: 杭州阿里云飞天信息技术有限公司(**HANGZHOU ALICLOUD FEITIAN INFORMATION TECHNOLOGY CO., LTD.**) [CN/CN]; 中国浙江省杭州市余杭区五常街道文一西路969号3幢5层553室, Zhejiang 310023 (CN)。

(72) 发明人: 李鹤 (**LI, He**); 中国浙江省杭州市西湖区三墩镇灯彩街1008号云谷园区, Zhejiang 310030 (CN)。

(74) 代理人:北京同钧律师事务所(**BEIJING TONGJUN LAW FIRM**); 中国北京市海淀区西直门北大街32号院1号楼7层808室, Beijing 100082 (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,

WO 2024/169529 A1

(54) **Title:** KNOWLEDGE BASE CONSTRUCTION METHOD, DATA RETRIEVAL METHOD AND APPARATUS, AND CLOUD DEVICE

(54) 发明名称: 知识库的构建方法、数据检索方法、装置和云设备

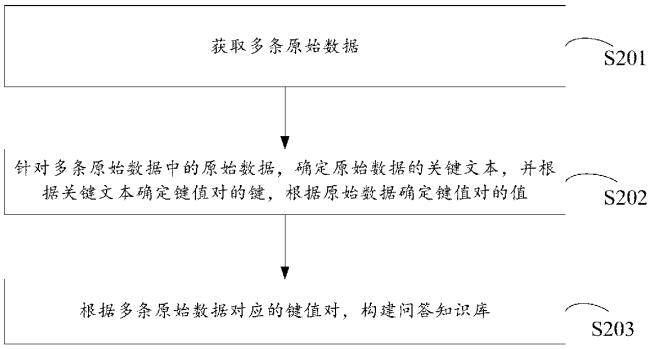

图 2

S201 Obtain a plurality of pieces of original data

S202 For original data in the plurality of pieces of original data, determine a key text of the original data, determine the key of a key-value pair according to the key text, and determine the value of the key-value pair according to the original data

S203 According to the key-value pairs corresponding to the plurality of pieces of original data, construct a question and answer knowledge base

(57) **Abstract:** The present application provides a knowledge base construction method, a data retrieval method and apparatus, and a cloud device. The knowledge base construction method comprises: obtaining a plurality of pieces of original data, the modalities of the plurality of pieces of original data being at least two of: text, image or video; for original data in the plurality of pieces of original data, determining a key text of the original data, determining the key of a key-value pair according to the key text, and determining the value of the key-value pair according to the original data, wherein the key text is used for describing the content of the original data; according to the key-value pairs corresponding to the plurality of pieces of original data, constructing a question and answer knowledge

[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)。

根据细则4.17的声明:
— 关于申请人有权要求在先申请的优先权(细则4.17(iii))

本国际公布:
— 包括国际检索报告(条约第21条(3))。

base, thereby allowing for constructing a question and answer knowledge base supporting different modalities of answer retrieval.

(57) 摘要: 本申请提供一种知识库的构建方法、数据检索方法、装置和云设备，该知识库的构建方法包括：获取多条原始数据，多条原始数据的模态为文本、图像或视频中的至少两种；针对多条原始数据中的原始数据，确定原始数据的关键文本，并根据关键文本确定键值对的键，根据原始数据确定键值对的值，关键文本用于描述原始数据的内容；根据多条原始数据对应的键值对，构建问答知识库，能够构建一个支持不同模态的答案检索的问答知识库。