

(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 2022/228127 A1

(43) 国际公布日
2022 年 11 月 3 日 (03.11.2022)

(51) 国际专利分类号:
G06F 16/35 (2019.01)

(21) 国际申请号: PCT/CN2022/086637

(22) 国际申请日: 2022 年 4 月 13 日 (13.04.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110476637.8 2021 年 4 月 29 日 (29.04.2021) CN

(71) 申请人: 京东科技控股股份有限公司(JINGDONG TECHNOLOGY HOLDING 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 221 室, Beijing 100176 (CN)。

(72) 发明人: 梁嘉辉(LINAG, Jiahui); 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 221 室, Beijing 100176 (CN)。 鲍军威(BAO, Junwei); 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 221 室, Beijing 100176 (CN)。

(74) 代理人: 北京清亦华知识产权代理事务所(普通合伙)(TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区北洼路 45 号 1 号楼 2 层 201, Beijing 100142 (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, JM, 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, 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,

(54) Title: ELEMENT TEXT PROCESSING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 要素文本处理方法、装置、电子设备和存储介质

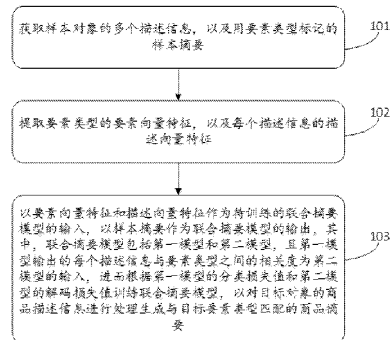


图 1

(57) Abstract: Provided are an element text processing method and apparatus, an electronic device, and a storage medium. The method comprises: obtaining a plurality of pieces of description information of a sample object and a sample abstract marked with an element type (101); extracting an element vector feature of the element type and a description vector feature of each description information (102); and taking the element vector feature and the description vector feature as input of a joint abstract model to be trained, and taking the sample abstract as output of the joint abstract model, wherein the joint abstract model comprises a first model and a second model, the relevance of each description information with the element type output by the first model is input of the second model, and thus the joint abstract model is trained according to a classification loss value of the first model and a decoding loss value of the second model, so as to process commodity description information of a target object to generate a commodity abstract matching a target element type (103).

- 101 Obtain a plurality of pieces of description information of a sample object and a sample abstract marked with an element type
- 102 Extract an element vector feature of the element type and a description vector feature of each description information
- 103 Take the element vector feature and the description vector feature as input of a joint abstract model to be trained, and take the sample abstract as output of the joint abstract model, wherein the joint abstract model comprises a first model and a second model, the relevance of each description information with the element type output by the first model is input of the second model, and thus the joint abstract model is trained according to a classification loss value of the first model and a decoding loss value of the second model, so as to process commodity description information of a target object to generate a commodity abstract matching a target element type

[见续页]

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：一种要素文本处理方法、装置、电子设备和存储介质。其中方法包括：获取样本对象的多个描述信息和用要素类型标记的样本摘要(101)；提取要素类型的要素向量特征和每个描述信息的描述向量特征(102)；以要素向量特征和描述向量特征作为待训练的联合摘要模型的输入，以样本摘要作为联合摘要模型的输出，其中联合摘要模型包括第一模型和第二模型，且第一模型输出的每个描述信息与要素类型之间的相关度为第二模型的输入，进而根据第一模型的分​​类损失值和第二模型的解码损失值训练联合摘要模型，以对目标对象的商品描述信息进行处理生成与目标要素类型匹配的商品摘要(103)。