

# (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  
2023 年 6 月 22 日 (22.06.2023)



(10) 国际公布号  
**WO 2023/108985 A1**

- (51) 国际专利分类号:  
**G06Q 40/06** (2012.01) **G06F 40/30** (2020.01)  
**G06K 9/62** (2022.01) **G06F 16/33** (2019.01)
- (21) 国际申请号: PCT/CN2022/090023
- (22) 国际申请日: 2022 年 4 月 28 日 (28.04.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
202111538838.2 2021 年 12 月 15 日 (15.12.2021) CN
- (71) 申请人: 平安科技 (深圳) 有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];  
中国广东省深圳市福田区福田街道福

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518048 (CN)。

(72) 发明人: 诸世卓 (ZHU, Shizhuo); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518048 (CN)。邵熹 (SHAO, Xi); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518048 (CN)。闻心远 (WEN, Xinyuan); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518048 (CN)。

(74) 代理人: 北京辰权知识产权代理有限公司 (BEIJING CHENQUAN INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区

(54) Title: METHOD FOR RECOGNIZING PROPORTION OF GREEN ASSET AND RELATED PRODUCT

(54) 发明名称: 绿色资产的占比的识别方法及相关产品

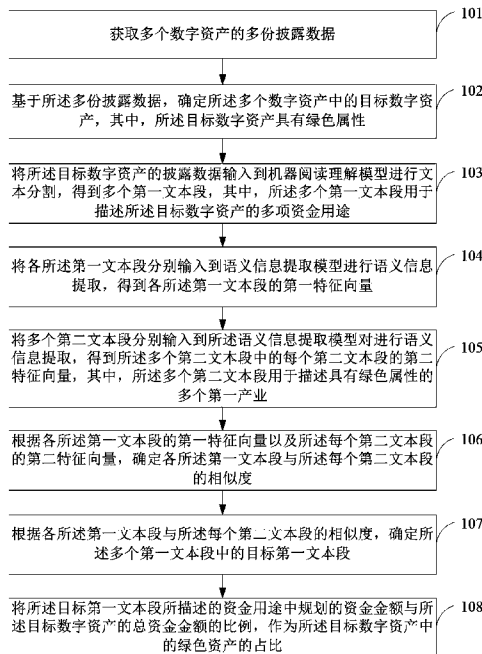


图 1

- 101 Acquire multiple pieces of disclosure data of multiple digital assets
- 102 Determine a target digital asset among the multiple digital assets on the basis of the multiple pieces of disclosure data, wherein the target digital asset has a green attribute
- 103 Input disclosure data of the target digital asset into a machine reading comprehension model for text segmentation to obtain multiple first text segments, wherein the multiple first text segments are used for describing multiple fund uses of the target digital asset
- 104 Input each of the first text segments into a semantic information extraction model respectively for semantic information extraction to obtain a first feature vector of each of the first text segments
- 105 Input multiple second text segments into the semantic information extraction model respectively for semantic information extraction to obtain a second feature vector of each of the multiple second text segments, wherein the multiple second text segments are used for describing multiple primary industries having green attributes
- 106 Determine the similarity between each of the first text segments and each of the second text segments according to the first feature vector of each of the first text segments and the second feature vector of each of the second text segments
- 107 Determine a target first text segment among the multiple first text segments according to the similarity between each of the first text segments and each of the second text segments
- 108 Use the proportion of the fund amount described in the target first text segment and planned in the fund uses in the total fund amount of the target digital asset as the proportion of a green asset in the target digital asset

(57) Abstract: A method for recognizing the proportion of a green asset and a related product, relating to the technical field of artificial intelligence. The method comprises: acquiring multiple pieces of disclosure data of multiple digital assets; determining a target digital asset on the basis of the multiple pieces of disclosure data; inputting disclosure data of the target digital asset into a machine reading comprehension model to obtain multiple first text segments; inputting each first text segment into a semantic information extraction model respectively to obtain a first feature vector of each first text segment; inputting multiple second text segments into the semantic

中关村东路 66 号世纪科贸大厦 C 座 2601  
室, Beijing 100190 (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, 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))。

information extraction model respectively to obtain a second feature vector of each second text segment; determining a target first text segment according to the first feature vector of each first text segment and the second feature vector of each second text segment; and using the proportion of the fund amount described in the target first text segment in the total fund amount of the target digital asset as the proportion of a green asset in the target digital asset.

(57) 摘要: 一种绿色资产的占比的识别方法及相关产品, 涉及人工智能技术领域, 该方法包括获取多个数字资产的多份披露数据; 基于多份披露数据, 确定目标数字资产; 将目标数字资产的披露数据输入到机器阅读理解模型, 得到多个第一文本段; 将各第一文本段分别输入到语义信息提取模型, 得到各第一文本段的第一特征向量; 将多个第二文本段分别输入到语义信息提取模型, 得到每个第二文本段的第二特征向量; 根据各第一文本段的第一特征向量以及每个第二文本段的第二特征向量, 确定目标第一文本段; 将目标第一文本段所描述的资金金额与目标数字资产的总资金金额的比例, 作为目标数字资产中的绿色资产的占比。

## 绿色资产的占比的识别方法及相关产品

### 优先权申明

本申请要求于 2021 年 12 月 15 日提交中国专利局、申请号为 202111538838.2，发明名称为“绿色资产的占比的识别方法及相关产品”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

### 技术领域

本申请涉及人工智能技术领域，具体涉及一种绿色资产的占比的识别方法及相关产品。

### 背景技术

在全球气候变化合作的大背景下，各个管理部门需要厘清在自己管辖范围内的绿色和非绿色资产规模，以便更加科学的部署碳达峰和碳中和的实现路径。

投资机构在实现碳达峰和碳中和的过程中扮演着非常重要的角色，其投资标的的选择实际上将引导企业向绿色产业和碳中和达标的方向发展。

发明人意识到投资机构在统计其绿色投资比例时，由于监管和保密的需要，不能进行跨部门共享，都是由各个部门进行人工统计，主观性较强，精度低。

### 发明内容

本申请实施例提供了一种绿色资产的占比的识别方法及相关产品，提高数字资产中的绿色资产的占比的识别精度。

第一方面，本申请实施例提供一种基于相似度的数字资产中的绿色资产的占比的识别方法，该方法包括：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述多个第一产业，所述多个第一产业为具有绿色属性的产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

第二方面，本申请实施例提供一种绿色资产的占比的识别装置，包括：获取单元和处理单元；

所述获取单元，用于获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

所述处理单元，用于基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

第三方面，本申请实施例提供一种电子设备，其中，包括：处理器和存储器，所述处理器与所述存储器相连，所述存储器用于存储计算机程序，所述处理器用于执行所述存储器中存储的计算机程序，所述计算机程序包括用于执行以下步骤的指令：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

第四方面，本申请实施例提供一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以使得计算机执行以下步骤的指令：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

第五方面，本申请实施例提供一种计算机程序产品，所述计算机程序产品包括存储了计算机程序的非瞬时性计算机可读存储介质，所述计算机可操作来使计算机执行如第一方面所述的方法。

实施本申请实施例，具有如下有益效果：

可以看出，在本申请实施例中，获取债券的披露数据，即可从多个债券中筛选出具有绿色属性的债券，即目标数字资产；然后，基于目标数字资产的披露数据，通过文本分割和相似度匹配技术自动识别出该目标数字资产中的绿色资产的占比，从而无需人工统计，减少了人力成本，以及避免了人工统计的主观性，提高了数字资产中的绿色资产的占比的识别精度。

## 附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例提供了一种基于相似度的数字资产中的绿色资产的占比的识别方法的流程示意图；

图 2 为本申请实施例提供的另一种基于相似度的数字资产中的绿色资产的占比的识别方法的流程示意图；

图 3 为本申请实施例提供了一种绿色资产的占比的识别装置的功能单元组成框图；

图 4 为本申请实施例提供了一种电子设备的结构示意图。

## 具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中，人工智能 (Artificial Intelligence, AI) 是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

首先说明，本申请的应用场景主要是识别出债券中的绿色资产的占比。因此，本申请实施例提供涉及的多个数字资产为投资机构所投资的多个债券，后续所提到的数字资产均可以理解为债券，比如，后续提到的第一数字资产，就是一个债券，不再区分，该投资机构为市场上的任意一个投资机构。另外，后续提到目标数字资产可以理解为具有绿色属性的债券。数字资产的所属企业，即是债券的发行企业。每个数字资产的披露数据，即为每个债券的披露数据，其中，每个债券的披露数据一般包含债券的名称、债券所属的企业、债券的规模、债券的各项资金以及各项资金的用途，等等。

参阅图 1，图 1 为本申请实施例提供的一种基于相似度的数字资产中的绿色资产的占比的识别方法的流程示意图。该方法应用于绿色资产的占比的识别装置。该方法包括以下步骤内容：

101：获取多个数字资产的多份披露数据。

示例性的，绿色资产的占比的识别装置可通过爬虫技术获取该多个数字资产的多份披露数据，其中，该多份披露数据与该多个数字资产一一对应。

102：基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性。

首先说明，该目标数字资产的数量可以为一个或多个，也就是说，该多个数字资产具有绿色属性的数字资产可以为一个或多个，本申请中主要以一个目标数字资产为例进行说明。

示例性的，根据第一数字资产的披露数据，确定第一数字资产的资产名称，即债券名称，

其中，该第一数字资产为该多个数字资产中的任意一个；然后，对该第一数字资产的资产名称进行关键词识别，得到第一关键词，其中，该第一关键词的数量为一个或多个；最后，若该第一关键词为预设关键词集中的关键词，则确定该第一数字资产为该目标数字资产。该预设关键词集合是由各个具有绿色属性且与数字资产相关的关键词构成的集合，即对各个绿色债券的债券名称进行关键词提取所得到的关键词构成的集合，比如，该预设关键词集合可以包括：“绿色债券”、“碳中和”，“节能”，等等。即从债券名称确定每个债券是否具有绿色属性，也就是确定每个债券是否为绿色债券。

示例性的，根据第一数字资产的披露数据，确定该第一数字资产的所属企业，即从披露数据中识别出该债券的发行企业；然后，确定该所属企业的所属行业，比如，可以将该所属企业的主营业务产品所属的行业，作为该所属企业的所属行业。最后，确定该所属行业是否为预设行业集合中的行业，若是，则确定该第一数字资产为目标数字资产，其中，该预设行业集合是由各个具有绿色属性的行业组成的集合。具体的，可获取预设文档，比如，《绿色债券支持项目目录》，然后对该预设文档进行实体提取，可得到一个或多个与绿色相关的绿色行业，例如，公共交通、污水处理等；然后，将这些绿色行业组成集合得到该预设行业集合。即从债券的所属行业确定出债券是否为绿色债券。

举例来说，若第一数字资产的披露数据为：债券的类型为“广州地铁集团有限公司 2020 年度第二期超短期融资券”，则从该披露数据中确定该债券的发行公司为广州地铁集团有限公司，且该发行公司的所属行业为公共交通。由于公共交通为预设行业集合中的行业，则确定该第一数字资产为目标数字资产。

示例性的，对第一数字资产的披露数据进行文本识别，从该披露数据中识别出第三文本段，其中，第三文本段为第一数字资产的披露数据中描述该第一数字资产的多项资金用途的文本段。即通过文本定位找到披露数据中描述该债券的各项资金用途的文本段，然后将各项资金用途的文本段从披露数据中提取出来，得到第三文本段；进一步的，对第三文本段进行语义信息提取，得到该第三文本段的第四特征向量；然后，根据该第四特征向量预测该第一数字资产具有绿色属性的概率；若该概率大于第二阈值，则确定该第一数字资产为目标数字资产。

在本申请的一个实施方式中，上述确定第一数字资产是否具有绿色属性的方式可以通过训练好的模型实现，该模型可以为 fasttext，textCNN，BERT 模型，等等，本申请对此不做限定。具体的，从债券样本中提取出用于描述资金用途的文本，将提取出的文本作为样本，并为该样本添加标签，该标签用于标识该债券样本是否具有绿色属性。应理解，在选择债券样

本时，应该分别选择具有绿色属性和非绿色属性的债券样本，以保证构造出的样本中包含有正样本和负样本；然后，基于提取出的样本以及该样本的标签进行模型训练，得到一个用于预测债券是否有绿色属性的预测模型；最后，通过该预测模型对第三文本段进行语义信息提取，得到该第三文本段的第四特征向量，并通过该预测模型对该第四特征向量进行处理，预测出该第一数字资产具有绿色属性的概率。

应说明，在实际应用中，可以优先选择债券名称或者债券的所属行业确定债券是否具有绿色属性，当这两种方式都无法确定时，则再通过模型预测的方式，去预测债券是否具有绿色属性。

103：将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途。

示例性的，预先训练好机器阅读理解（Machine Reading Comprehension，MRC）模型，然后将该目标数字资产的披露数据输入 MRC 模型中进行文本分割，得到多个第一文本段。

具体的，首先设定 MRC 所要解决的问题为“哪些文本是用来描述资金的用途的”，输入的文章为目标数字资产的披露数据；然后，通过 MRC 模型的编码层对问题进行编码，得到第一向量；通过 MRC 模型的编码层对披露数据中的各个文本段进行编码，得到与各个文本段对应的第二向量；然后，将第一向量和各个文本段的第二向量输入到 MRC 模型的交互层进行交互，得到问题和各个文本段之间的相似度，将相似度大于预设阈值的文本段作为该多个第一文本段。

举例来说，通过 MRC 模型对目标数字资产的披露数据进行文本分割，可得到如表 1 所示的多个第一文本段。

表 1：

| 总资金额<br>( 亿元 ) | 第一文本段                               | 项目<br>资金 | 项目资金百<br>分比 ( % ) | 资金用途                  |
|----------------|-------------------------------------|----------|-------------------|-----------------------|
| 5              | 所筹资金 2.40 亿元<br>用于雅砻江两河口水<br>电站项目建设 | 2.4      | 48                | 雅砻江两河口<br>水电站项目建<br>设 |
| 5              | 0.36 亿元用于金沙<br>江银江水电站项目建<br>设       | 0.36     | 7.2               | 金沙江银江水<br>电站项目建<br>设  |
| 5              | 0.24 亿元用于雅砻<br>江卡拉水电站项目建            | 0.24     | 4.8               | 雅砻江卡拉水<br>电站项目建<br>设  |

|   |                        |   |    |        |
|---|------------------------|---|----|--------|
|   | 设                      |   |    |        |
| 5 | 剩余 2.00 亿元用于<br>补充营运资金 | 2 | 40 | 补充营运资金 |

104：将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量。

其中，该语义信息提取模型是预先训练好的。下面描述对该语义信息提取模型的训练过程。

示例性的，首先构建训练样本。例如，从多个债券的披露数据中提取出与资金用途相关的文本段，并为每个文本段打上标签，其中，该标签用于标识该文本段所描述的资金用途具有绿色属性的实际情况，其中，该资金用途可以为用于绿色产业或者为非绿色产业。例如，表 1 中示出的资金用途：“用于雅奢江卡拉水电站项目建设”用于的产业项目为“雅奢江卡拉水电站项目建设”，则该资金用途具有绿色属性，即该资金用于用于绿色产业；然后，将打上有标签的各个文本段作为训练样本。进一步的，构建初始模型，其中，该初始模型可以为 Bert 模型，其包括语义信息提取模型和多层感知器（Multilayer Perceptron，MLP），其中，该语义信息提取模型和多层感知器的模型参数都是随机初始化得到；然后将训练样本输入到该语义信息提取模型进行语义信息提取，得到训练样本的第三特征向量；将该第三特征向量输入到多层感知器，得到该训练样本属于具有绿色属性的产业的概率；最后，根据该训练样本属于具有绿色属性的产业的概率，以及该训练样本的标签，对该初始模型进行训练，即对语义信息提取模型以及多层感知器的模型参数进行调整，得到目标模型，将该目标模型中的多层感知器删除，得到语义信息提取模型。

示例性的，可以将每个第一文本段输入到语义信息提取模型进行语义信息提取，得到每个第一文本段的第一特征向量。

在实际应用中，在得到目标模型之后，也可以不对目标模型进行删除，直接保留整个目标模型；然后，将每个第一文本段输入到目标模型中进行概率预测，得到每个第一文本段描述的资金用途属于绿色产业的概率，若该概率大于概率阈值，则确定该第一文本段为目标第一文本段，不需要进行相似度的计算，即可直接确定出目标第一文本段，提高绿色资产的占

比的识别效率。

105：将多个第二文本段分别输入到所述语义信息提取模型对进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业。

示例性的，获取多个具有绿色属性的产业，即绿色产业。具体的，对《绿色债券支持项目目录》PDF 文档进行实体（该实体为产业）识别，得到多个产业，将该多个产业作为该多个第一产业，并从该 PDF 文档中提取出用于描述该多个第一产业的多个第二文本段，其中，多个第二文本段用于该多个第一产业；同样的，将每个第二文本段输入到上述的语义信息提取模型进行语义信息提取，得到每个第二文本段的第二特征向量。

106：根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度。

示例性的，可以确定每个第一文本段的第一特征向量与每个第二文本段的第二特征向量之间的相似度，比如，该相似度可以通过两个特征向量之间的欧式距离表征，并将两个特征向量之间的相似度作为每个第一文本段与每个第二文本段之间的相似度。

107：根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段。

示例性的，根据每个第一文本段与每个第二文本段的相似度，从多个第二文本段中确定出与每个第一文本段对应的最大相似度，若该最大相似度大于阈值，则将该第一文本段作为目标文本段。具体的，若该最大相似度大于阈值，则说明该第一文本段所描述的资金用途所属的产业为该最大相似度对应的第二文本段描述的第一产业，即该资金用途所应用的产业是一个绿色产业，因此，可以确定出该资金用途具有绿色属性。

108：将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

示例性的，获取该目标第一文本段所描述的资金用途中所规划的资金金额，并获取该目标数字资产的总资金金额，即获取该目标债券的总规模；然后，将该目标第一文本段所描述的资金用途中所规划的资金金额与总资金金额的比例，作为该目标数字资产中的绿色资产的占比。

应说明，该目标第一文本段的数量为一个或多个，也就是说，该目标数字资产的多个资金用途中有多项资金用途所应用的产业具有绿色属性。则可以对每个目标第一文本段所描述的资金用途中规划的资金金额与目标数字资产的总资金金额的比例，作为每个目标第一文本

段的绿色比例；然后，对每个目标第一文本段的绿色比例求和，得到目标数字资产中的绿色资产的占比。

可以看出，在本申请实施例中，获取债券的披露数据，即可从多个债券中筛选出具有绿色属性的债券，即目标数字资产；然后，基于目标数字资产的披露数据，通过文本分割和相似度匹配技术自动识别出该目标数字资产中的绿色资产的占比（即债券的绿色比例），从而无需人工统计，减少了人力成本，以及人工统计的主观性，提高了数字资产中的绿色资产的占比的识别精度。

参阅图 2，图 2 为本申请实施例提供的另一种相似度的数字资产中的绿色资产的占比的识别方法的流程示意图。该实施例中与图 1 所示的实施例相同的内容，此处不再重复描述。本实施例的方法包括以下步骤：

201：获取  $t$  时刻下投资机构所投资的多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应，所述  $t$  时刻为任意时刻。

示例性的，基于爬虫技术以及时间戳，获取多个数字资产在  $t$  时刻下的多份披露数据。

202：基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性。

203：将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途。

204：将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量。

205：将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业。

206：根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度。

207：根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段。

208：将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

209：获取所述  $t$  时刻下所述目标数字资产的净值。

210：获取所述 t 时刻下所述投资机构持有所述目标数字资产的份额。

211：根据所述目标数字资产的净值、份额以及绿色资产的占比，确定所述 t 时刻下所述投资机构持有所述目标数字资产的绿色规模。

示例性的，绿色规模可以通过公式 (1) 表示：

$$S_i = s_i * V_i * R_i \quad \text{公式 (1)}；$$

其中， $s_i$  为所述目标数字资产的绿色资产的占比， $V_i$  为所述 t 时刻下所述目标数字资产的净值， $R_i$  为所述 t 时刻下所述投资机构持有所述目标数字资产的份额。

可选的，当目标数字资产的数量为多个时，也就是说，该投资机构所投资的多个数字资产中有多个绿色数字资产时，可以按照上述图 2 示出的方法确定投资机构持有目标数字资产的绿色规模，然后对持有每个目标数字资产的绿色规模进行求和，得到该投资机构持有多个目标数字资产的绿色规模。

示例性的，投资机构持有多个目标数字资产的绿色规模可以通过公式 (2) 表示：

$$S_i = \sum_{i=1}^n s_i * V_i * R_i \quad \text{公式 (2)}；$$

其中， $n$  为多个目标数字资产的数量， $s_i$  为  $n$  个目标数字资产中的第  $i$  个目标数字资产的绿色资产的占比， $V_i$  为 t 时刻下第  $i$  个目标数字资产的净值， $R_i$  为所述 t 时刻下所述投资机构持有第  $i$  个目标数字资产的份额。

可以看出，在本申请实施例中，获取债券的披露数据，即可从多个债券中筛选出具有绿色属性的债券，即目标数字资产；然后，基于目标数字资产的披露数据，通过文本分割和相似度匹配技术自动识别出该目标数字资产中的绿色资产的占比，从而无需人工统计，减少了人力成本，以及人工统计的主观性，提高了数字资产的绿色资产的占比的识别精度。由于统计出的绿色资产的占比比较精确，从而统计出的该投资机构所持有的绿色债券的规模的精度也比较高，因此可以精确的引导企业向绿色产业和碳中和达标的方向发展。

参阅图 3，图 3 本申请实施例提供的一种绿色资产的占比的识别装置的功能单元组成框图。绿色资产的占比的识别装置 300 包括：获取单元 301 和处理单元 302；

获取单元 301，用于获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

处理单元 302，用于基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一

文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

在一些可能的实施方式中，将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量之前，处理单元 302，还用于：

构建初始模型，所述初始模型包括语义信息提取模型和多层感知器；

将训练样本输入到所述语义信息提取模型进行语义信息提取，得到所述训练样本的第三特征向量；

将所述第三特征向量输入到所述多层感知器，得到所述训练样本属于具有绿色属性的产业的概率；

根据所述训练样本属于具有绿色属性的产业的概率，以及所述训练样本的标签，对所述初始模型进行训练，得到目标模型；

将所述目标模型中的所述多层感知器删除，得到所述语义信息提取模型。

在一些可能的实施方式中，在根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段方面，处理单元 302，具体用于：

根据各所述第一文本段与所述每个第二文本段的相似度，确定各所述第一文本段对应的最大相似度；

将所述多个第一文本段中所对应的最大相似度大于第一阈值的第一文本段，作为所述目标第一文本段。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理单元 302，具体用于：

根据第一数字资产的披露数据，确定所述第一数字资产的资产名称，所述第一数字资产为所述多个数字资产中的任意一个；

对所述资产名称进行关键词提取，得到第一关键词；

若所述第一关键词属于预设关键词集中的关键词，确定所述第一数字资产为所述目标数字资产，其中，所述预设关键词集合是由各个具有绿色属性，且与数字资产相关的关键词构成的集合。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理单元 302，具体用于：

根据第一数字资产的披露数据，确定所述第一数字资产的所属企业，所述第一数字资产为所述多个数字资产中的任意一个；

确定所述所属企业的所属行业；

若所述所属行业为预设行业集合中的行业，则确定所述第一数字资产为所述目标数字资产，其中，所述预设行业集合是由各个具有绿色属性的行业组成的集合。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理单元 302，具体用于：

对第一数字资产的披露数据进行文本识别，从所述披露数据中识别出第三文本段，其中，所述第三文本段为所述第一数字资产的披露数据中描述所述第一数字资产的多项资金用途所组成的文本段；

对所述第三文本段进行语义信息提取，得到所述第三文本段的第四特征向量；

根据所述第四特征向量预测所述第一数字资产具有绿色属性的概率；

若所述第一数字资产具有绿色属性的概率大于第二阈值，则将所述第一数字资产作为所述目标数字资产。

在一些可能的实施方式中，所述多个数字资产为  $t$  时刻下投资机构所投资的数字资产，所述  $t$  时刻为任意时刻；获取单元 301，还用于获取所述  $t$  时刻下所述目标数字资产的净值；获取所述  $t$  时刻下所述投资机构持有所述目标数字资产的份额；

处理单元 302，还用于根据所述目标数字资产的净值、份额以及绿色资产的占比，确定所述  $t$  时刻下所述投资机构持有所述目标数字资产的绿色规模；

所述绿色规模满足以下公式：

$$S_i = s_i * V_i * R_i$$

$s_i$  为所述目标数字资产的绿色资产的占比， $V_i$  为所述  $t$  时刻下所述目标数字资产的净值，

$R_i$ 为所述  $t$  时刻下所述投资机构持有所述目标数字资产的份额。

参阅图 4，图 4 为本申请实施例提供的一种电子设备的结构示意图。如图 4 所示，电子设备 400 包括收发器 401、处理器 402 和存储器 403。它们之间通过总线 404 连接。存储器 403 用于存储计算机程序和数据，并可以将存储器 403 存储的数据传输给处理器 402。

处理器 402 用于读取存储器 403 中的计算机程序执行以下操作：

控制收发器 401 获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

在一些可能的实施方式中，将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量之前，处理器 402 还用于执行以下操作：

构建初始模型，所述初始模型包括语义信息提取模型和多层感知器；

将训练样本输入到所述语义信息提取模型进行语义信息提取，得到所述训练样本的第三特征向量；

将所述第三特征向量输入到所述多层感知器，得到所述训练样本属于具有绿色属性的产业的概率；

根据所述训练样本属于具有绿色属性的产业的概率，以及所述训练样本的标签，对所述初始模型进行训练，得到目标模型；

将所述目标模型中的所述多层感知器删除，得到所述语义信息提取模型。

在一些可能的实施方式中，在根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段方面，处理器 402 具体用于执行以下操作：

根据各所述第一文本段与所述每个第二文本段的相似度，确定各所述第一文本段对应的最大相似度；

将所述多个第一文本段中所对应的最大相似度大于第一阈值的第一文本段，作为所述目标第一文本段。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理器 402 具体用于执行以下操作：

根据第一数字资产的披露数据，确定所述第一数字资产的资产名称，所述第一数字资产为所述多个数字资产中的任意一个；

对所述资产名称进行关键词提取，得到第一关键词；

若所述第一关键词属于预设关键词集合中的关键词，确定所述第一数字资产为所述目标数字资产，其中，所述预设关键词集合是由各个具有绿色属性，且与数字资产相关的关键词构成的集合。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理器 402 具体用于执行以下操作：

根据第一数字资产的披露数据，确定所述第一数字资产的所属企业，所述第一数字资产为所述多个数字资产中的任意一个；

确定所述所属企业的所属行业；

若所述所属行业为预设行业集合中的行业，则确定所述第一数字资产为所述目标数字资产，其中，所述预设行业集合是由各个具有绿色属性的行业组成的集合。

在一些可能的实施方式中，在基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产方面，处理器 402 具体用于执行以下操作：

对第一数字资产的披露数据进行文本识别，从所述披露数据中识别出第三文本段，其中，所述第三文本段为所述第一数字资产的披露数据中描述所述第一数字资产的多项资金用途所组成的文本段；

对所述第三文本段进行语义信息提取，得到所述第三文本段的第四特征向量；

根据所述第四特征向量预测所述第一数字资产具有绿色属性的概率；

若所述第一数字资产具有绿色属性的概率大于第二阈值，则将所述第一数字资产作为所述目标数字资产。

在一些可能的实施方式中，所述多个数字资产为  $t$  时刻下投资机构所投资的数字资产，所述  $t$  时刻为任意时刻；处理器 402 还用于执行以下操作：

控制收发器 401 获取所述  $t$  时刻下所述目标数字资产的净值；获取所述  $t$  时刻下所述投资机构持有所述目标数字资产的份额；

根据所述目标数字资产的净值、份额以及绿色资产的占比，确定所述  $t$  时刻下所述投资机构持有所述目标数字资产的绿色规模；

所述绿色规模满足以下公式：

$$S_t = s_t * V_t * R_t$$

$s_t$  为所述目标数字资产的绿色资产的占比， $V_t$  为所述  $t$  时刻下所述目标数字资产的净值， $R_t$  为所述  $t$  时刻下所述投资机构持有所述目标数字资产的份额。

具体地，上述收发器 401 可为图 3 所述的实施例的绿色资产的占比的识别装置 300 的获取单元 301，上述处理器 402 可以为图 4 所述的实施例的绿色绿色资产的占比的识别装置 300 的处理单元 302。

应理解，本申请中的电子设备可以包括智能手机（如 Android 手机、iOS 手机、Windows Phone 手机等）、平板电脑、掌上电脑、笔记本电脑、移动互联网设备 MID（Mobile Internet Devices，简称：MID）或穿戴式设备等。上述电子设备仅是举例，而非穷举，包括但不限于上述电子设备。在实际应用中，上述电子设备还可以包括：智能车载终端、计算机设备等等。

本申请实施例还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以实现如上述方法实施例中记载的任何一种基于相似度的数字资产的绿色资产的占比的识别方法的部分或全部步骤。所述计算机可读存储介质可以是非易失性，也可以是易失性。

本申请实施例还提供一种计算机程序产品，所述计算机程序产品包括存储了计算机程序的非瞬时性计算机可读存储介质，所述计算机程序可操作来使计算机执行如上述方法实施例中记载的任何一种基于相似度的数字资产的绿色资产的占比的识别方法的部分或全部步骤。

以上对本申请实施例进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对

于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

## 权 利 要 求 书

1.一种基于相似度的数字资产中的绿色资产的占比的识别方法，其中，包括：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

2.根据权利要求1所述的方法，其中，将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量之前，所述方法还包括：

构建初始模型，所述初始模型包括语义信息提取模型和多层感知器；

将训练样本输入到所述语义信息提取模型进行语义信息提取，得到所述训练样本的第三特征向量；

将所述第三特征向量输入到所述多层感知器，得到所述训练样本属于具有绿色属性的产业的概率；

根据所述训练样本属于具有绿色属性的产业的概率，以及所述训练样本的标签，对所述初始模型进行训练，得到目标模型；

将所述目标模型中的所述多层感知器删除，得到所述语义信息提取模型。

3.根据权利要求1或2所述的方法，其中，所述根据各所述第一文本段与所述每个第二

文本段的相似度，确定所述多个第一文本段中的目标第一文本段，包括：

根据各所述第一文本段与所述每个第二文本段的相似度，确定各所述第一文本段对应的最大相似度；

将所述多个第一文本段中所对应的最大相似度大于第一阈值的第一文本段，作为所述目标第一文本段。

4.根据权利要求 1-3 中任一项所述的方法，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的资产名称，所述第一数字资产为所述多个数字资产中的任意一个；

对所述资产名称进行关键词提取，得到第一关键词；

若所述第一关键词属于预设关键词集合中的关键词，确定所述第一数字资产为所述目标数字资产，其中，所述预设关键词集合是由各个具有绿色属性，且与数字资产相关的关键词构成的集合。

5.根据权利要求 1-3 中任一项所述的方法，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的所属企业，所述第一数字资产为所述多个数字资产中的任意一个；

确定所述所属企业的所属行业；

若所述所属行业为预设行业集合中的行业，则确定所述第一数字资产为所述目标数字资产，其中，所述预设行业集合是由各个具有绿色属性的行业组成的集合。

6.根据权利要求 1-3 中任一项所述的方法，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

对第一数字资产的披露数据进行文本识别，从所述披露数据中识别出第三文本段，其中，所述第三文本段为所述第一数字资产的披露数据中描述所述第一数字资产的多项资金用途所组成的文本段；

对所述第三文本段进行语义信息提取，得到所述第三文本段的第四特征向量；

根据所述第四特征向量预测所述第一数字资产具有绿色属性的概率；

若所述第一数字资产具有绿色属性的概率大于第二阈值，则将所述第一数字资产作为所述目标数字资产。

7.根据权利要求 1-6 中任一项所述的方法，其中，所述多个数字资产为 t 时刻下投资机构

所投资的数字资产，所述 t 时刻为任意时刻；所述方法还包括：

获取所述 t 时刻下所述目标数字资产的净值；

获取所述 t 时刻下所述投资机构持有所述目标数字资产的份额；

根据所述目标数字资产的净值、份额以及绿色资产的占比，确定所述 t 时刻下所述投资机构持有所述目标数字资产的绿色规模；

所述绿色规模满足以下公式：

$$S_i = s_i * V_i * R_i$$

$s_i$  为所述目标数字资产的绿色资产的占比， $V_i$  为所述 t 时刻下所述目标数字资产的净值， $R_i$  为所述 t 时刻下所述投资机构持有所述目标数字资产的份额。

8. 一种绿色资产的占比的识别装置，其中，包括：获取单元和处理单元；

所述获取单元，用于获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

所述处理单元，用于基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

9. 一种电子设备，其中，包括：处理器和存储器，所述处理器与所述存储器相连，所述存储器用于存储计算机程序，所述处理器用于执行所述存储器中存储的计算机程序，所述计算机程序包括用于执行以下步骤的指令：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

10.根据权利要求9所述的电子设备，其中，将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量之前，所述步骤还包括：

构建初始模型，所述初始模型包括语义信息提取模型和多层感知器；

将训练样本输入到所述语义信息提取模型进行语义信息提取，得到所述训练样本的第三特征向量；

将所述第三特征向量输入到所述多层感知器，得到所述训练样本属于具有绿色属性的产业的概率；

根据所述训练样本属于具有绿色属性的产业的概率，以及所述训练样本的标签，对所述初始模型进行训练，得到目标模型；

将所述目标模型中的所述多层感知器删除，得到所述语义信息提取模型。

11.根据权利要求10所述的电子设备，其中，所述根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段，包括：

根据各所述第一文本段与所述每个第二文本段的相似度，确定各所述第一文本段对应的

最大相似度；

将所述多个第一文本段中所对应的最大相似度大于第一阈值的第一文本段，作为所述目标第一文本段。

12.根据权利要求 11 所述的电子设备，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的资产名称，所述第一数字资产为所述多个数字资产中的任意一个；

对所述资产名称进行关键词提取，得到第一关键词；

若所述第一关键词属于预设关键词集合中的关键词，确定所述第一数字资产为所述目标数字资产，其中，所述预设关键词集合是由各个具有绿色属性，且与数字资产相关的关键词构成的集合。

13.根据权利要求 11 所述的电子设备，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的所属企业，所述第一数字资产为所述多个数字资产中的任意一个；

确定所述所属企业的所属行业；

若所述所属行业为预设行业集合中的行业，则确定所述第一数字资产为所述目标数字资产，其中，所述预设行业集合是由各个具有绿色属性的行业组成的集合。

14.根据权利要求 11 所述的电子设备，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

对第一数字资产的披露数据进行文本识别，从所述披露数据中识别出第三文本段，其中，所述第三文本段为所述第一数字资产的披露数据中描述所述第一数字资产的多项资金用途所组成的文本段；

对所述第三文本段进行语义信息提取，得到所述第三文本段的第四特征向量；

根据所述第四特征向量预测所述第一数字资产具有绿色属性的概率；

若所述第一数字资产具有绿色属性的概率大于第二阈值，则将所述第一数字资产作为所述目标数字资产。

15.一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以使得计算机执行以下步骤的指令：

获取多个数字资产的多份披露数据，其中，所述多个数字资产与所述多份披露数据一一

对应；

基于所述多份披露数据，确定所述多个数字资产中的目标数字资产，其中，所述目标数字资产具有绿色属性；

将所述目标数字资产的披露数据输入到机器阅读理解模型进行文本分割，得到多个第一文本段，其中，所述多个第一文本段用于描述所述目标数字资产的多项资金用途；

将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量；

将多个第二文本段分别输入到所述语义信息提取模型进行语义信息提取，得到所述多个第二文本段中的每个第二文本段的第二特征向量，其中，所述多个第二文本段用于描述具有绿色属性的多个第一产业；

根据各所述第一文本段的第一特征向量以及所述每个第二文本段的第二特征向量，确定各所述第一文本段与所述每个第二文本段的相似度；

根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段；

将所述目标第一文本段所描述的资金用途中规划的资金金额与所述目标数字资产的总资金金额的比例，作为所述目标数字资产中的绿色资产的占比。

16.根据权利要求 15 所述的计算机可读存储介质，其中，将各所述第一文本段分别输入到语义信息提取模型进行语义信息提取，得到各所述第一文本段的第一特征向量之前，所述步骤还包括：

构建初始模型，所述初始模型包括语义信息提取模型和多层感知器；

将训练样本输入到所述语义信息提取模型进行语义信息提取，得到所述训练样本的第三特征向量；

将所述第三特征向量输入到所述多层感知器，得到所述训练样本属于具有绿色属性的产业的概率；

根据所述训练样本属于具有绿色属性的产业的概率，以及所述训练样本的标签，对所述初始模型进行训练，得到目标模型；

将所述目标模型中的所述多层感知器删除，得到所述语义信息提取模型。

17.根据权利要求 16 所述的计算机可读存储介质，其中，所述根据各所述第一文本段与所述每个第二文本段的相似度，确定所述多个第一文本段中的目标第一文本段，包括：

根据各所述第一文本段与所述每个第二文本段的相似度，确定各所述第一文本段对应的

最大相似度；

将所述多个第一文本段中所对应的最大相似度大于第一阈值的第一文本段，作为所述目标第一文本段。

18.根据权利要求 17 所述的计算机可读存储介质，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的资产名称，所述第一数字资产为所述多个数字资产中的任意一个；

对所述资产名称进行关键词提取，得到第一关键词；

若所述第一关键词属于预设关键词集合中的关键词，确定所述第一数字资产为所述目标数字资产，其中，所述预设关键词集合是由各个具有绿色属性，且与数字资产相关的关键词构成的集合。

19.根据权利要求 17 所述的计算机可读存储介质，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

根据第一数字资产的披露数据，确定所述第一数字资产的所属企业，所述第一数字资产为所述多个数字资产中的任意一个；

确定所述所属企业的所属行业；

若所述所属行业为预设行业集合中的行业，则确定所述第一数字资产为所述目标数字资产，其中，所述预设行业集合是由各个具有绿色属性的行业组成的集合。

20.根据权利要求 17 所述的计算机可读存储介质，其中，所述基于所述多个数字资产的披露数据，确定所述多个数字资产中的目标数字资产，包括：

对第一数字资产的披露数据进行文本识别，从所述披露数据中识别出第三文本段，其中，所述第三文本段为所述第一数字资产的披露数据中描述所述第一数字资产的多项资金用途所组成的文本段；

对所述第三文本段进行语义信息提取，得到所述第三文本段的第四特征向量；

根据所述第四特征向量预测所述第一数字资产具有绿色属性的概率；

若所述第一数字资产具有绿色属性的概率大于第二阈值，则将所述第一数字资产作为所述目标数字资产。

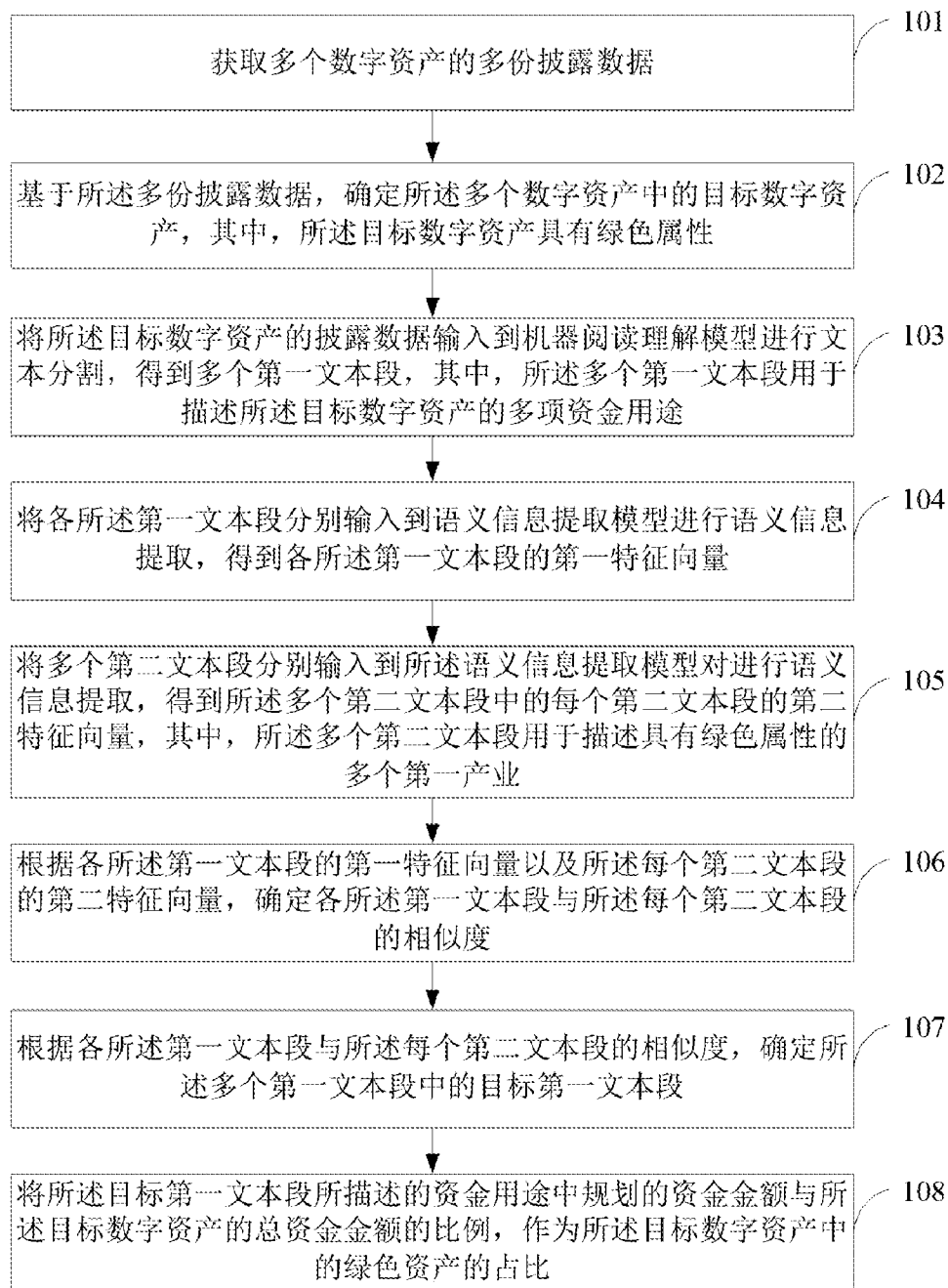


图 1

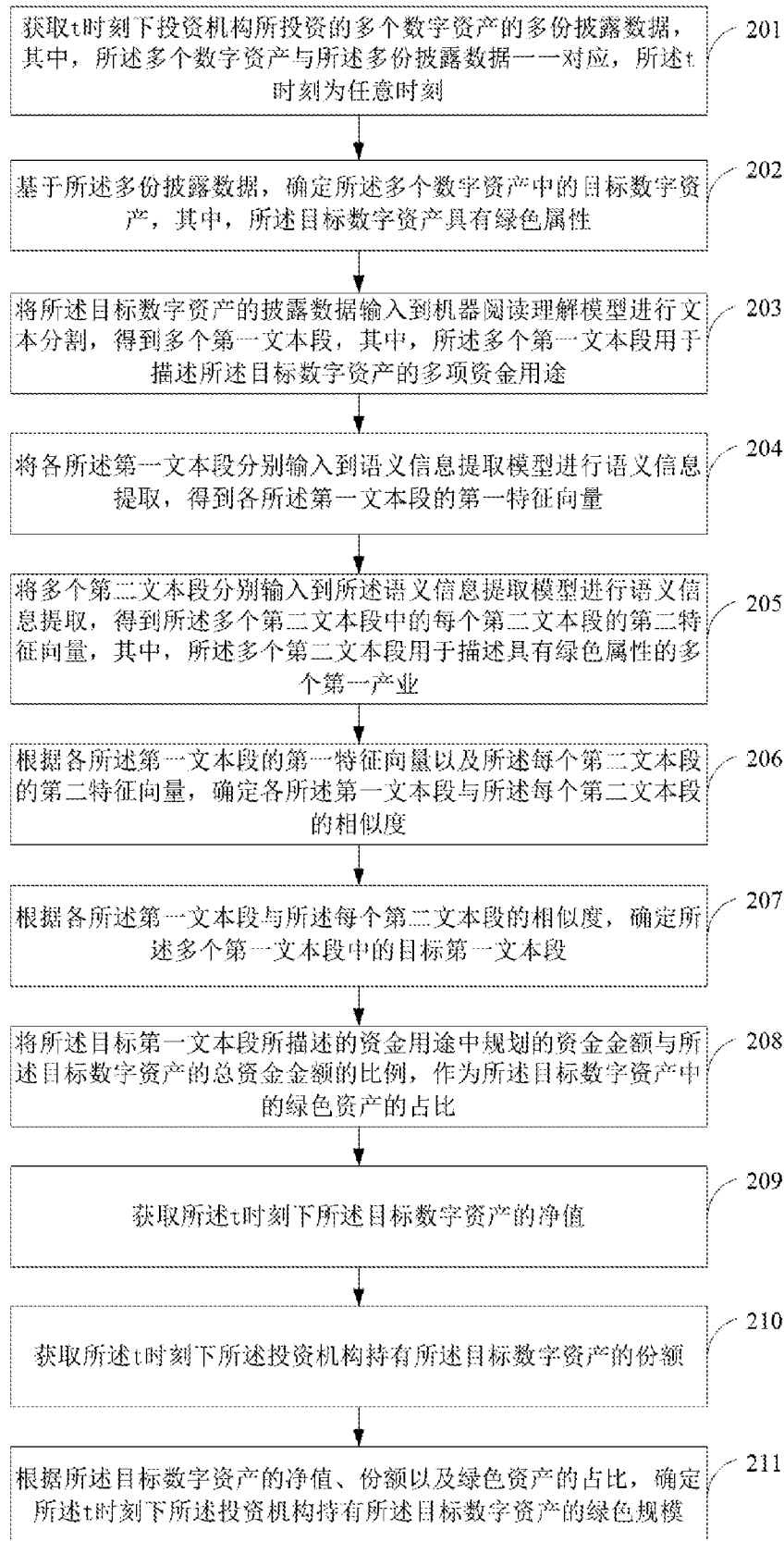


图 2

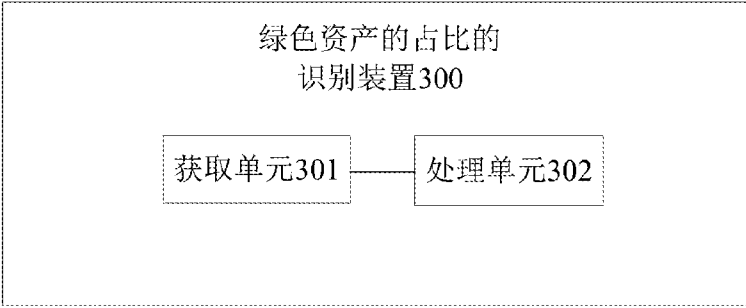


图 3

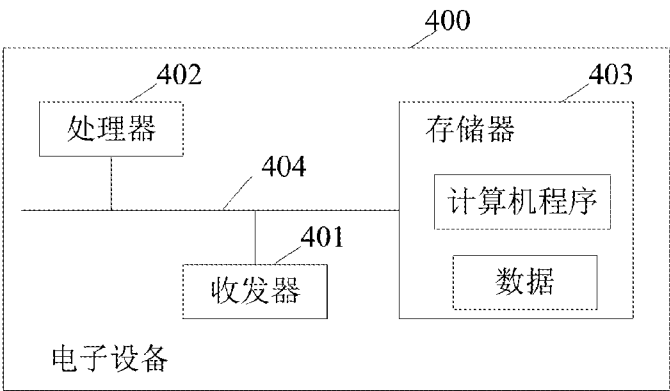


图 4

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/090023

**A. CLASSIFICATION OF SUBJECT MATTER**

G06Q 40/06(2012.01)i; G06K 9/62(2022.01)i; G06F 40/30(2020.01)i; G06F 16/33(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06Q; G06K; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, WPI, CNPAT, CNKI, IEEE: 文本, 切分, 切割, 多, 特征向量, 相似度, 匹配, 绿色, 投资, 基金, 资产, 金额, 占比, 比例, 识别, 投资书, 披露数据, text, file, divide, slice, cut, multi, feature vector, similarity, match, green, investment, found, asset, amount, fraction, ratio, recognition, investment book, disclosure data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                  | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| PX        | CN 114240672 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 25 March 2022 (2022-03-25)<br>claims 1-10                                                  | 1-20                  |
| A         | CN 111427996 A (UNISOUND INTELLIGENT TECHNOLOGY CO., LTD. et al.) 17 July 2020 (2020-07-17)<br>abstract, and description, paragraphs 4-14 and 41-42 | 1-20                  |
| A         | CN 113705192 A (PING AN BANK CO., LTD.) 26 November 2021 (2021-11-26)<br>entire document                                                            | 1-20                  |
| A         | CN 113658007 A (SHANGHAI PUDONG DEVELOPMENT BANK CO., LTD.) 16 November 2021 (2021-11-16)<br>entire document                                        | 1-20                  |
| A         | CN 112183989 A (ZHONGCAI LVZHI BEIJING INFORMATION CONSULTING CO., LTD.) 05 January 2021 (2021-01-05)<br>entire document                            | 1-20                  |
| A         | US 2019005020 A1 (ELSEVIER, INC.) 03 January 2019 (2019-01-03)<br>entire document                                                                   | 1-20                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

17 August 2022

Date of mailing of the international search report

29 August 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing  
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2022/090023**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 114240672  | A  | 25 March 2022                        | None                    |            |    |                                      |
| CN                                        | 111427996  | A  | 17 July 2020                         | None                    |            |    |                                      |
| CN                                        | 113705192  | A  | 26 November 2021                     | None                    |            |    |                                      |
| CN                                        | 113658007  | A  | 16 November 2021                     | None                    |            |    |                                      |
| CN                                        | 112183989  | A  | 05 January 2021                      | None                    |            |    |                                      |
| US                                        | 2019005020 | A1 | 03 January 2019                      | WO                      | 2019006115 | A1 | 03 January 2019                      |

## 国际检索报告

国际申请号

PCT/CN2022/090023

|                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                     |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A. 主题的分类</b><br>G06Q 40/06(2012.01)i; G06K 9/62(2022.01)i; G06F 40/30(2020.01)i; G06F 16/33(2019.01)i<br><br>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                     |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| <b>B. 检索领域</b><br>检索的最低限度文献(标明分类系统和分类号)<br>G06Q; G06K; G06F<br><br>包含在检索领域中的除最低限度文献以外的检索文献<br><br>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))<br>EP0DOC, WPI, CNPAT, CNKI, IEEE: 文本, 切分, 切割, 多, 特征向量, 相似度, 匹配, 绿色, 投资, 基金, 资产, 金额, 占比, 比例, 识别, 投资书, 披露数据, text, file, divide, slice, cut, multi, feature vector, similarity, match, green, investment, found, asset, amount, fraction, ratio, recognition, investment book, disclosure data               |                                                                                                                                                                                     |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| <b>C. 相关文件</b>                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| 类 型*                                                                                                                                                                                                                                                                                                                                                                                                                                               | 引用文件, 必要时, 指明相关段落                                                                                                                                                                   | 相关的权利要求                              |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| PX                                                                                                                                                                                                                                                                                                                                                                                                                                                 | CN 114240672 A (平安科技深圳有限公司) 2022年3月25日 (2022 - 03 - 25)<br>权利要求1-10                                                                                                                 | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                  | CN 111427996 A (云知声智能科技股份有限公司 等) 2020年7月17日 (2020 - 07 - 17)<br>摘要、说明书第4-14和41-42段                                                                                                  | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                  | CN 113705192 A (平安银行股份有限公司) 2021年11月26日 (2021 - 11 - 26)<br>全文                                                                                                                      | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                  | CN 113658007 A (上海浦东发展银行股份有限公司) 2021年11月16日 (2021 - 11 - 16)<br>全文                                                                                                                  | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                  | CN 112183989 A (中财绿指北京信息咨询有限公司) 2021年1月5日 (2021 - 01 - 05)<br>全文                                                                                                                    | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                  | US 2019005020 A1 (ELSEVIER, INC.) 2019年1月3日 (2019 - 01 - 03)<br>全文                                                                                                                  | 1-20                                 |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| <input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| <table><tr><td>* 引用文件的具体类型:<br/>“A” 认为不特别相关的表示了现有技术一般状态的文件<br/>“E” 在国际申请日的当天或之后公布的在先申请或专利<br/>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)<br/>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br/>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</td><td>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件<br/>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性<br/>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性<br/>“&amp;” 同族专利的文件</td></tr></table> |                                                                                                                                                                                     |                                      | * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件 | “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件<br>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性<br>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性<br>“&” 同族专利的文件 |
| * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件                                                                                                                                                                                                                                          | “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件<br>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性<br>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性<br>“&” 同族专利的文件 |                                      |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| 国际检索实际完成的日期<br>2022年8月17日                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                     | 国际检索报告邮寄日期<br>2022年8月29日             |                                                                                                                                                                                                           |                                                                                                                                                                                     |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                     | 受权官员<br>姜晓庆<br>电话号码 86-(10)-53961427 |                                                                                                                                                                                                           |                                                                                                                                                                                     |

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2022/090023

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 114240672  | A  | 2022年3月25日     | 无    |            |    |                |
| CN          | 111427996  | A  | 2020年7月17日     | 无    |            |    |                |
| CN          | 113705192  | A  | 2021年11月26日    | 无    |            |    |                |
| CN          | 113658007  | A  | 2021年11月16日    | 无    |            |    |                |
| CN          | 112183989  | A  | 2021年1月5日      | 无    |            |    |                |
| US          | 2019005020 | A1 | 2019年1月3日      | WO   | 2019006115 | A1 | 2019年1月3日      |