

(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
2021 年 5 月 20 日 (20.05.2021)



(10) 国际公布号
WO 2021/093643 A1

- (51) 国际专利分类号:
G06F 21/10 (2013.01) *G06K 9/62* (2006.01)
- (21) 国际申请号: PCT/CN2020/126232
- (22) 国际申请日: 2020 年 11 月 3 日 (03.11.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201911093190.5 2019 年 11 月 11 日 (11.11.2019) CN
- (71) 申请人: 深圳前海微众银行股份有限公司 (WEBANK CO., LTD) [CN/CN]; 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518000 (CN)。
- (72) 发明人: 蔡远航(CAI, Yuanhang); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518000 (CN)。 郑少杰(ZHENG, Shaojie); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518000

(CN)。 付勇(FU, Yong); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518000 (CN)。 范增虎(FAN, Zenghu); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518000 (CN)。

(74) 代理人: 深圳市世纪恒程知识产权代理事务所 (CENFO INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市南山区西丽街道松坪山社区松坪山路 3 号奥特迅电力大厦 201, Guangdong 518052 (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:** COPYRIGHT AUTHENTICATION METHOD, DEVICE, APPARATUS AND SYSTEM AND COMPUTER READ-ABLE STORAGE MEDIUM

(54) 发明名称: 版权认证方法、装置、设备、系统及计算机可读存储介质

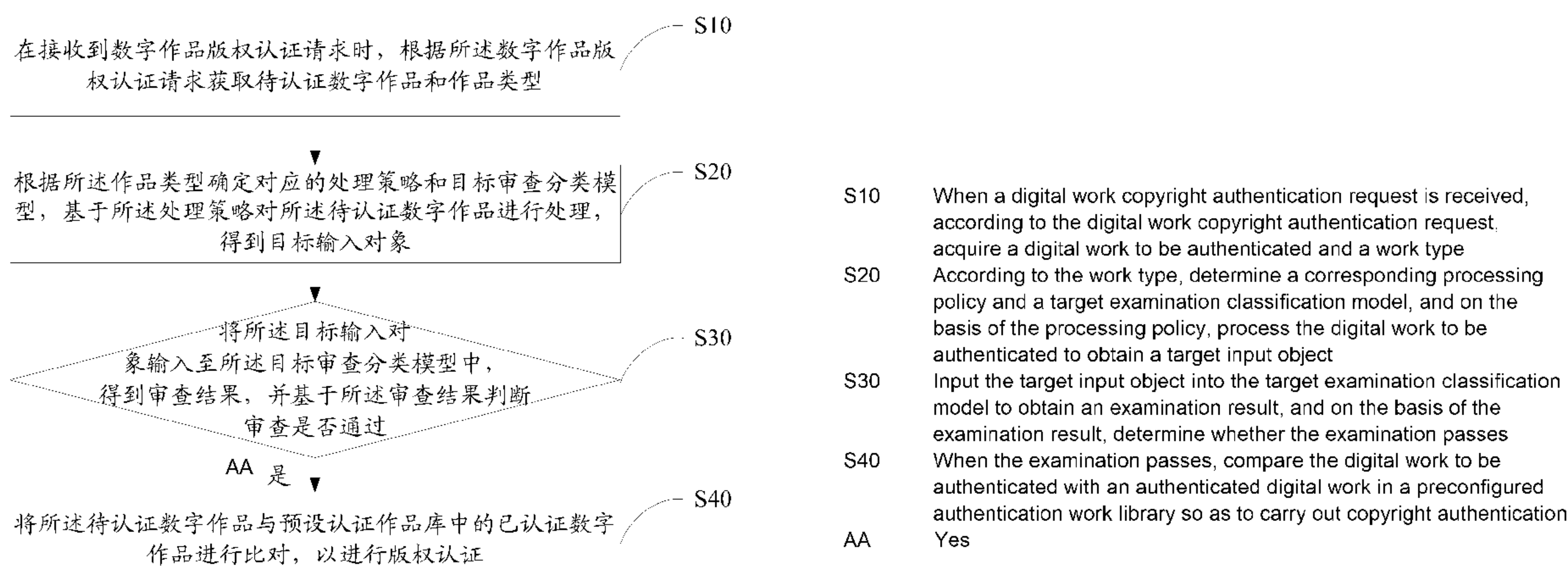


图 2

(57) **Abstract:** The present invention relates to the technical field of fintech and provides a copyright authentication method, device, apparatus and system and a computer readable storage medium. The copyright authentication method comprises: when a digital work copyright authentication request is received, according to the digital work copyright authentication request, acquiring a digital work to be authenticated and a work type (S10); according to the work type, determining a corresponding processing policy and a target examination classification model, and on the basis of the processing policy, processing the digital work to be authenticated to obtain a target input object (S20); inputting the target input object into the target examination classification model to obtain an examination result, and on the basis of the examination result, determining whether the examination passes (S30); and when the examination passes, comparing the digital work to be authenticated with an authenticated digital work in a preconfigured authentication work library so as to carry out copyright authentication (S40).

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：一种版权认证方法、装置、设备、系统及计算机可读存储介质，涉及金融科技技术领域。该版权认证方法包括：在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型 (S10)；根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象 (S20)；将所述目标输入对象输入至所述目标审查分类模型中，得到审查结果，并基于所述审查结果判断审查是否通过 (S30)；在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证 (S40)。

版权认证方法、装置、设备、系统及计算机可读存储介质

优先权信息

5 本申请要求于2019年11月11日申请的、申请号为201911093190.5、名称为“版权认证方法、装置、设备、系统及计算机可读存储介质”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

10 本申请涉及金融科技（Fintech）技术领域，尤其涉及一种版权认证方法、装置、设备、系统及计算机可读存储介质。

背景技术

15 随着计算机技术的发展，越来越多的技术（大数据、分布式、区块链 Blockchain、人工智能等）应用在金融领域，传统金融业正在逐步向金融科技（Fintech）转变，但由于金融行业的安全性、实时性要求，也对技术提出了更高的要求。

目前作品版权认证主要通过线下人工受理的方式完成，具体流程为：1）作品著作权人向知识产权代理人提交个人信息及作品；2）由知识产权代理人判断作品是否具备登记条件，判定版权登记类型；3）若作品符合版权申请条件，由知识产权代理人向版权中心提交作品登记申请表；4）版权中心接到申请后，对申请资料进行审核，并决定是否下发作品的著作权登记证书。整个流程大约需要 20~30 个工作日，周期较长。随着互联网技术的普及，网络上每天都会产生十万级的原创数字作品，线下人工受理作品版权认证的方式由于存在耗时耗力、成本高、效率低和周期长等缺陷，已经无法满足当下的需求，使得大量侵权盗版作品在网络上传播。因此，亟需一种数字作品的版权认证方法，以缩短版权认证周期、提高版权认证效率，及时对作品著作权人的版权进行保护。

20

25

发明内容

本申请的主要目的在于提供一种版权认证方法、装置、设备、系统及计算机可读存储介质，旨在缩短版权认证周期，提高版权认证效率。

30 为实现上述目的，本申请提供一种版权认证方法，所述版权认证方法包括：

在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型；

根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象；

35 将所述目标输入对象输入至所述目标审查分类模型中，得到审查结果，并基于所述审查结果判断审查是否通过；

在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。

40 在一实施例中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

若所述作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型；

基于所述第一处理策略对所述待认证数字作品进行切词处理，得到第一切词文本；

45 将所述第一切词文本输入预设词向量模型，得到所述第一切词文本中各切词的第一词向量；

根据所述第一词向量得到所述待认证数字作品对应的第一文档向量，其中，所述目标输入对象为所述第一文档向量。

在一实施例中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

5 若所述作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型；

基于所述第二处理策略对所述待认证数字作品进行预处理，得到输入图片，其中，所述预处理包括缩放处理和灰度处理，所述目标输入对象为所述输入图片。

10 在一实施例中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

若所述作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型；

基于所述第三处理策略将所述待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品；

15 对所述转换后的待认证数字作品进行切词处理，得到第二切词文本；

将所述第二切词文本输入预设词向量模型，得到所述第二切词文本中各切词的第二词向量；

根据所述第二词向量得到所述转换后的待认证数字作品对应的第二文档向量，其中，所述目标输入对象为所述第二文档向量。

20 在一实施例中，所述目标审查分类模型包括多个，所述审查结果的数量与所述目标审查分类模型的数量相同，所述基于所述审查结果判断审查是否通过的步骤包括：

检测多个所述审查结果是否均为审查合格；

若多个所述审查结果均为审查合格，则判定审查通过；

若多个所述审查结果中至少存在一个为审查不合格，则判定审查不通过。

25 在一实施例中，若所述作品类型为文字作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值；

30 根据所述第一相似度值从所述已认证文字作品中筛选得到第一预设数量的相似文字作品；

计算所述相似文字作品与所述待认证数字作品之间的第一最长公共子序列，并根据所述第一最长公共子序列的长度计算所述相似文字作品与所述待认证数字作品之间的长度比值，得到第一计算结果；

检测所述第一计算结果中是否存在大于第一预设阈值的长度比值；

35 若存在大于第一预设阈值的长度比值，则判定版权认证不通过；

若不存在大于第一预设阈值的长度比值，则判定版权认证通过。

在一实施例中，所述通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值的步骤包括：

通过预设文档搜索引擎对所述待认证数字作品进行分词处理，得到分词集；

40 通过所述预设文档搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引，并根据所述倒排索引结果计算所述分词集中各分词对应的分值；

对各分词的分值进行加和处理，得到所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。

45 在一实施例中，若所述作品类型为图片作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

通过预设图像检索引擎计算所述待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值；

根据所述第二相似度值从所述已认证图片作品中筛选得到第二预设数量的相似图片作品；

5 提取所述待认证数字作品的第一尺度不变特征变换 SIFT 特征向量，并提取所述相似图片作品的第二 SIFT 特征向量；

计算所述第一 SIFT 特征向量与所述第二 SIFT 特征向量之间的余弦距离，得到第二计算结果；

检测所述第二计算结果中是否存在大于第二预设阈值的余弦距离；

10 若存在大于第二预设阈值的余弦距离，则判定版权认证不通过；

若不存在大于第二预设阈值的余弦距离，则判定版权认证通过。

在一实施例中，若所述作品类型为音频作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

将所述待认证数字作品转换成文字作品类型，得到待认证音频文字作品；

15 通过预设文档检索引擎计算所述待认证音频文字作品与预设认证作品库中的已认证音频文字作品之间的第三相似度值；

根据所述第三相似度值从所述已认证音频文字作品中检索得到第三预设数量的相似音频文字作品；

20 计算所述相似音频文件作品与所述待认证音频文字作品之间的第二最长公共子序列，并根据所述第二最长公共子序列的长度计算所述相似音频文字作品与所述待认证音频文字作品之间的长度比值，得到第三计算结果；

检测所述第三计算结果中是否存在大于第三预设阈值的长度比值；

若存在大于第三预设阈值的长度比值，则判定版权认证不通过；

若不存在大于第三预设阈值的长度比值，则判定版权认证通过。

25 在一实施例中，所述版权认证方法还包括：

在版权认证通过时，获取所述待认证数字作品的作品信息；

基于所述作品信息生成对应的版权认证信息，并根据所述版权认证信息生成数据上链请求；

30 将所述数据上链请求发送至版权认证联盟链，以使得所述版权认证联盟链基于所述数据上链请求完成对所述待认证数字作品的上链操作。

此外，为实现上述目的，本申请还提供一种版权认证装置，所述版权认证装置包括：

第一获取模块，用于在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型；

35 处理模块，用于根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象；

审查模块，用于将所述目标输入对象输入至所述目标审查分类模型中，得到审查结果，并基于所述审查结果判断审查是否通过；

40 版权认证模块，用于在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。

此外，为实现上述目的，本申请还提供一种版权认证设备，所述版权认证设备包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的版权认证程序，所述版权认证程序被所述处理器执行时实现如上所述的版权认证方法的步骤。

45 此外，为实现上述目的，本申请还提供一种版权认证系统，所述版权认证系统包括版

权认证设备和版权认证联盟链；其中，

所述版权认证设备为如上所述的版权认证设备；

所述版权认证联盟链，用于接收所述版权认证设备发送的数据上链请求；

基于所述数据上链请求获取待上链数据信息，并基于共识算法完成对所述待上链数据信息的上链操作。

此外，为实现上述目的，本申请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上存储有版权认证程序，所述版权认证程序被处理器执行时实现如上所述的版权认证方法的步骤。

本申请提供一种版权认证方法、装置、设备、系统及计算机可读存储介质，在接收到数字作品版权认证请求时，根据数字作品版权认证请求获取待认证数字作品和作品类型；根据作品类型确定对应的处理策略和目标审查分类模型，基于处理策略对待认证数字作品进行处理，得到目标输入对象；将目标输入对象输入至目标审查分类模型中，得到审查结果，并基于审查结果判断审查是否通过；并在在审查通过时，将待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。通过上述方式，可实现对数字作品版权的在线认证，可以针对不同类型的数字作品在秒级的时间复杂度内完成版权认证，相比于现有技术中通过人工进行版权认证，本申请可降低人力成本、缩短版权认证周期、提高版权认证效率，从而可及时对作品著作权人的版权进行保护。

附图说明

图1为本申请实施例方案涉及的硬件运行环境的设备结构示意图；

图2为本申请版权认证方法第一实施例的流程示意图；

图3为本申请版权认证系统的系统结构示意图；

图4为本申请版权认证装置第一实施例的功能模块示意图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

参照图1，图1为本申请实施例方案涉及的硬件运行环境的设备结构示意图。

本申请实施例版权认证设备可以是智能手机，也可以是PC（Personal Computer，个人计算机）、平板电脑、便携计算机等终端设备。

如图1所示，该版权认证设备可以包括：处理器1001，例如CPU，通信总线1002，用户接口1003，网络接口1004，存储器1005。其中，通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏（Display）、输入单元比如键盘（Keyboard），可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口（如Wi-Fi接口）。存储器1005可以是高速RAM存储器，也可以是稳定的存储器（non-volatile memory），例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

本领域技术人员可以理解，图1中示出的版权认证设备结构并不构成对版权认证设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。

如图1所示，作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及版权认证程序。

在图1所示的终端中，网络接口1004主要用于连接后台服务器，与后台服务器进行

数据通信；用户接口 1003 主要用于连接客户端，与客户端进行数据通信；而处理器 1001 可以用于调用存储器 1005 中存储的版权认证程序，并执行以下版权认证方法的各个步骤。

基于上述硬件结构，提出本申请版权认证方法的各实施例。

5 本申请提供一种版权认证方法。

参照图 2，图 2 为本申请版权认证方法第一实施例的流程示意图。

在本实施例中，该版权认证方法包括：

步骤 S10，在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型；

10 在本实施例中，该版权认证方法应用于版权认证系统中，该版权认证系统包括版权认证设备和版权认证联盟链，其中，本实施例的版权认证方法是由版权认证设备实现的，该设备搭载有版权认证系统。版权认证联盟链可以由版权认证机构节点、公证机构节点、司法机构节点以及外联节点组成，用于接收版权认证设备发送的数据上链请求，进而基于数据上链请求获取待上链数据信息，然后基于共识算法完成对该待上链数据信息的上链操作，即实现基于区块链的版权认证。

在本实施例中，当用户需要对其作品进行版权认证时，可通过用户端（如 PC 个人计算机、智能手机等）的对应软件上传其数字作品（如文字作品、图片作品和音频作品等），并填写相关信息（包括但不限于作品类型、著作人信息等），进而触发数字作品版权认证请求，此时，版权认证系统在接收到数字作品版权认证请求时，根据数字作品版权认证请求获取待认证数字作品和作品类型。当然，可以理解，在具体实施时，用户在触发数字作品版权认证请求时，可以只上传其数字作品，版权认证系统可在获取到待认证数字作品后，可根据其格式判断其对应的作品类型。

步骤 S20，根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象；

25 在获取到待认证数字作品和作品类型后，根据作品类型确定对应的处理策略和目标审查分类模型，基于处理策略对待认证数字作品进行处理，得到目标输入对象。

具体的，若作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型；然后，基于第一处理策略对待认证数字作品进行切词处理，得到第一切词文本；将第一切词文本输入预设词向量模型，得到第一切词文本中各切词的第一词向量；进而根据第一词向量得到待认证数字作品对应的第一文档向量，其中，目标输入对象为第一文档向量。

若作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型；基于第二处理策略对待认证数字作品进行预处理，得到输入图片，其中，预处理包括缩放处理和灰度处理，目标输入对象为输入图片。

35 若作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型；基于第三处理策略将待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品；然后按与文字作品的处理方法进行处理，即，对转换后的待认证数字作品进行切词处理，得到第二切词文本；进而将第二切词文本输入预设词向量模型，得到第二切词文本中各切词的第二词向量；根据第二词向量得到转换后的待认证数字作品对应的第二文档向量，其中，目标输入对象为第二文档向量。

具体的执行过程可参照下述第二实施例，此处不作赘述。

步骤 S30，将所述目标输入对象输入至所述目标审查分类模型中，得到审查结果，并基于所述审查结果判断审查是否通过；

45 在得到目标输入对象之后，将目标输入对象输入至目标审查分类模型中，得到审查结果，并基于审查结果判断审查是否通过。其中，在审查时，主要是为了审查用户提交的作

品是否涉及恐怖宣传、邪教宣传、政治敏感以及黄、赌、毒，对应的目标审查分类模型可以包括 6 类。即，目标审查分类模型可以包括恐怖宣传类审查分类模型、邪教宣传类审查分类模型、政治敏感类审查分类模型、涉黄类审查分类模型、涉赌类审查分类模型、涉毒类审查分类模型。当目标审查分类模型可以包括多个时，审查结果的数量与目标审查分类模型的数量相同，即对应的审查结果也包括多个。在基于多个审查结果判断审查是否通过时，需检测多个审查结果是否均为审查合格；若多个所述审查结果均为审查合格，则说明待认证数字作品不涉及不良信息（恐怖宣传/邪教宣传/政治敏感/黄/赌/毒），此时判定审查通过；若多个所述审查结果中至少存在一个为审查不合格，则说明待认证数字作品涉及不良信息（恐怖宣传/邪教宣传/政治敏感/黄/赌/毒）中一种或多种，此时判定审查不通过。

当然，在具体实施例中，还可以针对各类型数字作品分别构建一个审查分类，对应的，审查结果只有一个，此时，则只需检测该审查结果是否为审查合格。

步骤 S40，在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。

在审查通过时，将待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。具体的，需针对不同类型的待认证数字作品采用不同的版权认证方法，具体的版权认证过程可参照下述第四实施例，此处不作赘述。通过将待认证数字作品与预设认证作品库中的已认证数字作品进行比对，可检测待认证作品与已认证作品之间是否存处在引用/抄袭关系，以确定是否对版本进行认证。

本申请实施例提供一种版权认证方法，在接收到数字作品版权认证请求时，根据数字作品版权认证请求获取待认证数字作品和作品类型；根据作品类型确定对应的处理策略和目标审查分类模型，基于处理策略对待认证数字作品进行处理，得到目标输入对象；将目标输入对象输入至目标审查分类模型中，得到审查结果，并基于审查结果判断审查是否通过；并在在审查通过时，将待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。通过上述方式，可实现对数字作品版权的在线认证，可以针对不同类型的数字作品在秒级的时间复杂度内完成版权认证，相比于现有技术中通过人工进行版权认证，本申请实施例可降低人力成本、缩短版权认证周期、提高版权认证效率，从而可及时对作品著作权人的版权进行保护。

进一步的，基于图 2 所示的第一实施例，提出本申请版权认证方法的第二实施例。

在本实施例中，作为其中一实施方式，步骤 S20 可以包括：

步骤 a11，若所述作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型；

步骤 a12，基于所述第一处理策略对所述待认证数字作品进行切词处理，得到第一切词文本；

步骤 a13，将所述第一切词文本输入预设词向量模型，得到所述第一切词文本中各切词的第一词向量；

步骤 a14，根据所述第一词向量得到所述待认证数字作品对应的第一文档向量，其中，所述目标输入对象为所述第一文档向量。

在本实施例中，对于文字作品的处理过程如下：

若作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型。其中，第一审查分类模型是预先训练好的，其类型可以为 SVM（Support Vector Machine，支持向量机）模型、贝叶斯模型、逻辑回归模型、卷积神经网络模型等二分类模型，下述训练过程以 SVM 模型进行说明。对于目标审查分类模型，以目标审查分类模型包括恐怖宣传类审查分类模型、邪教宣传类审查分类模型、政治敏感类审查分类模型、涉黄类审查分类模型、涉赌类审查分类模型、涉毒类审查分类模型这 6 类

为例进行说明，对应的，第一审查分类模型也包括 6 类，其训练过程为：分别取已标注的 5 万条涉及和 5 万条不涉及恐怖宣传的文字作品，对每个文字作品进行切词后，通过预设词向量模型（在一实施例中为 word2vec 模型）得到词向量后，对词向量按对应维度进行相加，得到各文字作品对应的文档向量，然后根据上述 10 万个文档向量训练一个 SVM 分类模型。接下来利用同样的方法，分别训练用于判断作品是否涉及邪教宣传、是否涉及政治敏感、是否涉黄、是否涉赌、是否涉毒的 5 个 SVM 分类模型。

在确定得到第一处理策略后，基于第一处理策略对待认证数字作品进行切词处理，得到第一切词文本，其中，切词处理可采用预设的工具，如中科院 NLPIR、哈工大 LTP、结巴分词等。具体的分词过程与现有技术相一致，此处不作赘述。

然后，将第一切词文本输入预设词向量模型，得到第一切词文本中各切词的第一词向量。其中，预设词向量模型在一实施例中为 word2vec（word to vector，用来产生词向量的相关模型），word2vec 将每一个中文词汇映射为一个高维向量（通常取 200 维向量），且对于任意两个中文词汇，语义上越相近，映射后得到的向量距离也越近。因此可以根据词向量之间的距离来描述中文词汇的语义相似性。

最后，根据第一词向量得到待认证数字作品对应的第一文档向量，其中，目标输入对象为第一文档向量，即后续将第一文档向量输入至对应的第一审查分类模型中，以得到审查结果。对于第一文档向量的获取，是对第一词向量按对应维度进行相加，即可得到对应的第一文档向量。

通过上述方式，可实现对文字作品类的数字作品进行处理，得到对应的目标输入对象，以便于后续输入至目标审查分类模型中，得到审查结果。

作为又一实施方式，步骤 S20 还可以包括：

步骤 a21，若所述作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型；

步骤 a22，基于所述第二处理策略对所述待认证数字作品进行预处理，得到输入图片，其中，所述预处理包括缩放处理和灰度处理，所述目标输入对象为所述输入图片。

在本实施例中，对于图片作品的处理过程如下：

若作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型。其中，第二审查分类模型是预先训练好的，其类型在一实施例中为基于卷积神经网络的分类模型。对于目标审查分类模型，以目标审查分类模型包括恐怖宣传类审查分类模型、邪教宣传类审查分类模型、政治敏感类审查分类模型、涉黄类审查分类模型、涉赌类审查分类模型、涉毒类审查分类模型这 6 类为例进行说明，对应的，第二审查分类模型也包括 6 类，其训练过程为：分别取已标注的 5 万条涉及和 5 万条不涉及恐怖宣传的图片作品，对每个图片作品进行预处理，预处理过程包括缩放处理和灰度处理，其中，缩放处理即为将图片的大小缩放为一预设尺寸，例如 128 像素*128 像素，灰度处理即将缩放后的图片转换为灰度图片，然后根据上述 10 万张预处理后的图片训练一个基于卷积神经网络的分类模型。接下来利用同样的方法，分别训练用于判断作品是否涉及邪教宣传、是否涉及政治敏感、是否涉黄、是否涉赌、是否涉毒的 5 个基于卷积神经网络的分类模型。

在确定得到第二处理策略后，基于第二处理策略对待认证数字作品进行预处理，得到输入图片，其中，预处理包括缩放处理和灰度处理，缩放处理即为将图片的大小缩放为一预设尺寸，例如 128 像素*128 像素，灰度处理即将缩放后的图片转换为灰度图片，目标输入对象为输入图片，即后续将输入图片输入至对应的第二审查分类模型中，以得到审查结果。

通过上述方式，可实现对图片作品类的数字作品进行处理，得到对应的目标输入对象，

以便于后续输入至目标审查分类模型中，得到审查结果。

作为另一实施方式，步骤 S20 还可以包括：

5 步骤 a31，若所述作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型；

步骤 a32，基于所述第三处理策略将所述待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品；

步骤 a33，对所述转换后的待认证数字作品进行切词处理，得到第二切词文本；

10 步骤 a34，将所述第二切词文本输入预设词向量模型，得到所述第二切词文本中各切词的第二词向量；

步骤 a35，根据所述第二词向量得到所述转换后的待认证数字作品对应的第二文档向量，其中，所述目标输入对象为所述第二文档向量。

在本实施例中，对于音频作品的处理过程如下：

15 若作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型。其中，第三审查分类模型是预先训练好的，其类型可以为 SVM（Support Vector Machine，支持向量机）模型、贝叶斯模型、逻辑回归模型、卷积神经网络模型等二分类模型，其中，第三审查分类模型可以与第一审查分类模型相同，也可以基于第一审查分类模型的训练方法训练得到的其他类型的二分类模型。

20 在确定得到第三处理策略后，先基于第三处理策略将待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品。具体的，可以通过语音识别工具将音频作品转换为文字作品类型。然后，对转换后的待认证数字作品进行切词处理，得到第二切词文本，其中，切词处理可采用预设的工具，如中科院 NLPIR、哈工大 LTP、结巴分词等。具体的分词过程与现有技术相一致，此处不作赘述。

25 然后，将第二切词文本输入预设词向量模型，得到第二切词文本中各切词的第二词向量。其中，预设词向量模型在一实施例中为 word2vec（word to vector，用来产生词向量的相关模型）。

30 最后，根据第二词向量得到待认证数字作品对应的第二文档向量，其中，目标输入对象为第二文档向量，即后续将第二文档向量输入值对应的第三审查分类模型中，以得到审查结果。对于第二文档向量的获取，是对第二词向量按对应维度进行相加，即可得到对应的第二文档向量。

通过上述方式，可实现对音频作品类的数字作品进行处理，得到对应的目标输入对象，以便于后续输入至目标审查分类模型中，得到审查结果。

基于上述第一实施例和第二实施例，提出本申请版权认证方法的第三实施例。

35 在本实施例中，所述目标审查分类模型包括多个，所述审查结果的数量与所述目标审查分类模型的数量相同，步骤“基于所述审查结果判断审查是否通过”包括：

步骤 b1，检测多个所述审查结果是否均为审查合格；

步骤 b2，若多个所述审查结果均为审查合格，则判定审查通过；

步骤 b3，若多个所述审查结果中至少存在一个为审查不合格，则判定审查不通过。

40 在本实施例中，目标审查分类模型可以包括多个，审查结果的数量与目标审查分类模型的数量相同，对应的审查结果也包括多个。对于审查是否通过的判断过程为：检测多个审查结果是否均为审查合格。

若多个所述审查结果均为审查合格，则说明待认证数字作品不涉及不良信息（恐怖宣传/邪教宣传/政治敏感/黄/赌/毒），此时判定审查通过；

45 若多个所述审查结果中至少存在一个为审查不合格，则说明待认证数字作品涉及不良信息（恐怖宣传/邪教宣传/政治敏感/黄/赌/毒）中一种或多种，此时判定审查不通过。

进一步的，基于上述第一实施例，提出本申请版权认证方法的第四实施例。

在本实施例中，若所述作品类型为文字作品，步骤 S40 包括：

5 步骤 c11，通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值；

步骤 c12，根据所述第一相似度值从所述已认证文字作品中筛选得到第一预设数量的相似文字作品；

10 步骤 c13，计算所述相似文字作品与所述待认证数字作品之间的第一最长公共子序列，并根据所述第一最长公共子序列的长度计算所述相似文字作品与所述待认证数字作品之间的长度比值，得到第一计算结果；

步骤 c14，检测所述第一计算结果中是否存在大于第一预设阈值的长度比值；

步骤 c15，若存在大于第一预设阈值的长度比值，则判定版权认证不通过；

步骤 c16，若不存在大于第一预设阈值的长度比值，则判定版权认证通过。

其中，步骤 c11 包括：

15 步骤 c111，通过预设文档搜索引擎对所述待认证数字作品进行分词处理，得到分词集；

步骤 c112，通过所述预设文档搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引，并根据所述倒排索引结果计算所述分词集中各分词对应的分值；

步骤 c113，对各分词的分值进行加和处理，得到所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。

20 在本实施例中，若作品类型为文字作品，其对应的版权认证过程如下：

通过预设文档搜索引擎计算待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。其中，预设文档搜索引擎在一实施例中为 ES (Elastic Search，弹性检索) 搜索引擎，ES 是一个分布式、高扩展、高实时的搜索与数据分析引擎，它能很方便的使大量数据具有搜索、分析和探索的能力。具体的，先通过 ES 搜索引擎对待认证数字作品进行分词处理，得到分词集，其中 ES 搜索引擎中自带分词器，可以对待认证数字作品进行分词。然后，通过 ES 搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引，并根据倒排索引结果计算分词集中各分词对应的分值。在通过 ES 搜索引擎对已认证文字作品进行倒排索引时，也是先通过 ES 搜索引擎自带的分词器进行分词，并会得到各分词的词频信息和位置信息，进而建立各分词与作品文档之间的倒排索引，其中倒排索引是一个字典类数据结构 (key-value)，字典的键(key)是一个个的分词，值(value)是包含该分词的作品列表，以及该分词在每个作品中的位置信息与词频信息，通过倒排索引，可以根据分词快速获取包含这个分词的文档列表及词频信息。在根据倒排索引结果计算分词集中各分词对应的分值时，可以根据倒排索引结果获取包含分词的文档、及分词在文档中的词频和逆文档频率，进而根据词频和逆文档频率计算各分词的分值。例如，对待认证数字作品进行分词后得到包含分词“ A”、“ B”、“ C”、“ D”、“ E” 和“ F” 的分词集，先基于已认证文字作品的倒排索引找到包含分词“ A” 的作品集，并计算分词“ A” 对应的分值，具体的，可以先计算作品集中每个作品的分值，每个作品的分值可以为作品中分词 A 的词频与词 A 的逆文档频率的乘积（当然也可以根据实际情况设定其他计算方式，如基于词频、逆文档频率和位置信息计算），进而对每个作品的分值进行加和得到分词“ A” 对应的分值；

40 在得到分词“ A” 对应的分值之后，对分词“ B”、“ C”、“ D”、“ E”、“ F” 执行相同的操作，以得到分词“ B”、“ C”、“ D”、“ E”、“ F” 分别对应的分值。此外，可以理解，预设认证作品库中的已认证文字作品，可预先采用 ES 搜索引擎按特定的索引结构进行存储，此时则无需执行“通过所述预设文档搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引”这一步骤。在得到的各分词对应的分值之后，对各分词的分值进行加和处理，得到待

45 认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。

在得到待认证数字作品与已认证文字作品之间的第一相似度值之后,根据第一相似度值从已认证文字作品中筛选得到第一预设数量的相似文字作品,其中,第一预设数量可以根据实际需要进行设定,例如,可设定为一预设数值,如 1000,还可以设定为第一相似度值大于一预设值的数量,此处不作具体限定。在筛选时,将第一相似度值按从大到小的顺序进行排序,取排名在前的第一预设数量的文字作品,作为相似文字作品。需要说明的是,考虑到两篇文字作品存在抄袭关系的前提是两篇作品的用词是一致的,且每个词语前后相邻的词语也是一致的。如果两篇作品的用词都不相同,那么这两篇新闻也一定不存在抄袭关系。因此,此处通过 ES 搜索引擎初步判断用词是否一致,以进行初步的筛选,其目的在于从已认证文字作品中筛选得到包含待认证文字作品中的全部或大部分切词的文字作品,以缩小认证比对的范围,节省服务器资源,进一步提高文字作品版权认证的效率。此外,需要说明的是,在具体实施时,当然也可以直接基于计算最长公共子序列的方式来计算待认证数字作品与预设认证作品库中的已认证文字作品之间的相似度值,进而基于该相似度值判定版权认证是否通过,但是相比而言,通过 ES 搜索引擎进行初筛的计算过程的复杂程度显然低于最长公共子序列的计算复杂程度,借助 ES 搜索引擎从千万级的已认证文字作品中检索出与待认证数字作品用词一致的相似度值为前 1000 作品集的耗时为毫秒级,因此,本实施例中通过 ES 搜索引擎先进行初筛,再与后续计算最长公共子序列相结合,相比于直接基于最长公共子序列计算相似度值的方式,可进一步提高作品版权认证效率。

然后,计算相似文字作品与待认证数字作品之间的第一最长公共子序列,其中,第一最长公共子序列中的“第一”并无实质含义,仅用于与后续的第二最长公共子序列进行区分。最长公共子序列,即找出两个序列的最长的公共子序列,例如,给定两个字符串 $X=\langle x_1, x_2, x_3, \dots, x_n \rangle$, $Y=\langle y_1, y_2, y_3, \dots, y_m \rangle$, 存在两个长度为 k 的下标序列 $\langle i_1, i_2, \dots, i_k \rangle$, $\langle j_1, j_2, \dots, j_k \rangle$, 使得字符串 X 和 Y 在对应下标 i 和 j 的位置的字符相等,满足上述要求的最长的下标序列对应的子字符串即为 X 和 Y 之间的最长公共子序列。如字符串 ABCDE 和字符串 XAYCDZ 之间的最长公共子序列为 ACD。然后,根据第一最长公共子序列的长度计算相似文字作品与待认证数字作品之间的长度比值,得到第一计算结果。第一最长公共子序列的数量对应的与相似文字作品的数量相同。对于长度比值的计算,例如,某一相似文字作品的第一最长公共子序列包括 100 个字符时,则其长度为 100,若待认证数字作品包括 1000 个字符,则长度比值为 $100/1000=0.1$ 。

然后,检测第一计算结果中是否存在大于第一预设阈值的长度比值,其中,第一预设阈值可根据实际需要设定,例如可设为 0.8,此处不作具体限定。若存在大于第一预设阈值的长度比值,则认为存在引用/抄袭关系,此时判定版权认证不通过;若不存在大于第一预设阈值的长度比值,则判定版权认证通过。

需要说明的是,在具体实施时,“计算相似文字作品与待认证数字作品之间的第一最长公共子序列,并根据第一最长公共子序列的长度计算相似文字作品与待认证数字作品之间的长度比值”与“检测是否存在大于第一预设阈值的长度比值”的过程可同时进行,即,依次计算各相似文字作品与待认证数字作品之间的第一最长公共子序列,然后根据第一最长公共子序列的长度计算相似文字作品与待认证数字作品之间的长度比值,进而该长度比值是否大于第一预设阈值,一旦检测到某一相似文字作品对应的长度比值大于第一预设阈值,则认为存在引用/抄袭关系,即可判定版权认证不通过,此时则无需计算其他相似文字作品与待认证数字作品之间的第一最长公共子序列及后续步骤,可节省服务器资源,进一步提高版权认证的效率。

通过上述方式,可检测待认证作品与已认证作品之间是否存处在引用/抄袭关系,以确定是否对版本进行认证。

在本实施例中，若所述作品类型为图片作品，步骤 S40 还可以包括：

步骤 c21，通过预设图像检索引擎计算所述待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值；

5 步骤 c22，根据所述第二相似度值从所述已认证图片作品中筛选得到第二预设数量的相似图片作品；

步骤 c23，提取所述待认证数字作品的第一尺度不变特征变换 SIFT 特征向量，并提取所述相似图片作品的第二 SIFT 特征向量；

步骤 c24，计算所述第一 SIFT 特征向量与所述第二 SIFT 特征向量之间的余弦距离，得到第二计算结果；

10 步骤 c25，检测所述第二计算结果中是否存在大于第二预设阈值的余弦距离；

步骤 c26，若存在大于第二预设阈值的余弦距离，则判定版权认证不通过；

步骤 c27，若不存在大于第二预设阈值的余弦距离，则判定版权认证通过。

在本实施例中，若所述作品类型为图片作品，其对应的版权认证过程如下：

15 通过预设图像检索引擎计算待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值。其中，预设图像检索引擎在一实施例中为 CBIR(Content-based image retrieval，基于内容的图像检索)引擎，CBIR 引擎的核心是使用图像的可视特征对图像进行检索。本质上讲，它是一种近似匹配技术，融合了计算机视觉、图像处理、图像理解和数据库等多个领域的技术成果，其中的特征提取和索引的建立可由计算机自动完成，避免了人工描述的主观性。用户检索的过程一般是提供一个样例图像(Queryby Example)或描绘一幅草图(Queryby Sketch)，系统抽取该查询图像的特征，然后与数据库中的特征进行比较，
20 并将与查询特征相似的图像返回给用户。需要说明的是，由于预设认证作品库中的已认证图片作品都是按预设尺寸存储的，故需对图片类的待认证数字作品进行对应的缩放处理，得到同样预设尺寸的待认证数字作品，再通过预设图像检索引擎计算缩放后的待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值。

25 在得到待认证数字作品与已认证图片作品之间的第二相似度值之后，根据第二相似度值从已认证图片作品中筛选得到第二预设数量的相似图片作品，其中，第二预设数量可以与第一预设数量相同或不同，可以根据实际需要进行设定，此处不作具体限定。在筛选时，将第二相似度值按从大到小的顺序进行排序，取排名在前的第二预设数量的文字作品，作为相似文字作品。需要说明的是，考虑到两个图片作品存在抄袭关系的前提是两个图片作品的颜色、形状、纹理等低层次特征是一致的，且各特征的排列方式也是一致的。如果两个
30 图片作品的低层次特征都不相似，那么这两个图片作品也一定不会存在抄袭关系。因此，此处通过 CBIR 引擎进行初步筛选的目的在于从已认证图片作品中筛选得到包含较多待认证图片作品中特征（如颜色特征、形状特征、纹理特征等）的图片作品，以缩小认证比对的范围，节省服务器资源，进一步提高图片作品版权认证的效率。当然，需要说明的是，
35 在具体实施时，也可以直接基于 SIFT 特征向量的方式来计算待认证数字作品与预设认证作品库中的已认证图片作品之间的相似度值，进而基于该相似度值判定版权认证是否通过，但是相比而言，通过 CBIR 引擎进行特征提取和相似度计算，相比于提取 SIFT 特征向量，其计算过程更为简单，效率更高，因此，本实施例中先通过 CBIR 引擎先进行初筛，再与后续基于 SIFT 特征向量计算相似度值相结合，相比于直接基于 SIFT 特征向量计算相似度值的方式，可进一步提高作品版权认证效率。
40

然后，分别提取待认证数字作品和相似图片作品的 SIFT (Scale-invariant feature transform，尺度不变特征变换)特征向量，即提取待认证数字作品的第一 SIFT 特征向量，并提取相似图片作品的第二 SIFT 特征向量。其中，SIFT 特征向量的提取过程为：在图像中检测出一系列的关键点，这些关键点对尺度缩放，旋转以及亮度变化无关，然后对上述
45 关键点分配梯度方向值，就可以得到一幅图像的 SIFT 特征向量。具体的提取过程与现有

技术相一致，此处不作具体描述。

进而，计算第一 SIFT 特征向量与第二 SIFT 特征向量之间的余弦距离，得到第二计算结果。本实施例中采用余弦距离表征两特征向量之间的相似度，在具体实施例中，还可以采用其他参数来表征，如欧式距离等。最后，检测第二计算结果中是否存在大于第二预设阈值的余弦距离；其中，第二预设阈值可以与第一预设阈值相同或不同，可根据实际需要设定，例如也可设为 0.8，此处不作具体限定。若存在大于第二预设阈值的余弦距离，则认为存在引用/抄袭关系，判定版权认证不通过；若不存在大于第二预设阈值的余弦距离，则判定版权认证通过。

同样的，在具体实施时，“提取待认证数字作品的第一 SIFT 特征向量，并提取相似图片作品的第二 SIFT 特征向量，然后计算第一 SIFT 特征向量与第二 SIFT 特征向量之间的余弦距离”与“检测是否存在大于第二预设阈值的余弦距离”的过程可同时进行，一旦检测到某一相似图片作品对应的余弦距离大于第二预设阈值，则认为存在引用/抄袭关系，即可判定版权认证不通过，此时则无需提取其他相似图片作品与待认证数字作品的 SIFT 特征向量，也无需计算对应的余弦距离及后续检测步骤，可节省服务器资源，进一步提高版权认证的效率。

在本实施例中，若所述作品类型为音频作品，步骤 S40 还可以包括：

步骤 c31，将所述待认证数字作品转换成文字作品类型，得到待认证音频文字作品；

步骤 c32，通过预设文档搜索引擎计算所述待认证音频文字作品与预设认证作品库中的已认证音频文字作品之间的第三相似度值；

步骤 c33，根据所述第三相似度值从所述已认证音频文字作品中检索得到第三预设数量的相似音频文字作品；

步骤 c34，计算所述相似音频文件作品与所述待认证音频文字作品之间的第二最长公共子序列，并根据所述第二最长公共子序列的长度计算所述相似音频文字作品与所述待认证音频文字作品之间的长度比值，得到第三计算结果；

步骤 c35，检测所述第三计算结果中是否存在大于第三预设阈值的长度比值；

步骤 c36，若存在大于第三预设阈值的长度比值，则判定版权认证不通过；

步骤 c37，若不存在大于第三预设阈值的长度比值，则判定版权认证通过。

在本实施例中，若作品类型为音频作品，其对应的版权认证过程如下：

将所述待认证数字作品转换成文字作品类型，得到待认证音频文字作品，然后，通过预设文档搜索引擎计算待认证数字作品与预设认证作品库中的已认证音频文字作品之间的第三相似度值。其中，预设文档搜索引擎在一实施例中为 ES 搜索引擎。预设认证作品库中的已认证音频文字作品，是基于语音识别工具将音频作品转换为文字作品类型得到的，可预先采用 ES 搜索引擎按特定的索引结构进行存储，以便于后续进行搜索。

在得到待认证数字作品与已认证音频文字作品之间的第三相似度值之后，根据第三相似度值从已认证音频文字作品中筛选得到第三预设数量的相似文字作品，其中，第三预设数量可以与第一预设数量和第二预设数量相同或不同，可根据实际需要设定，此处不作具体限定。在筛选时，将第三相似度值按从大到小的顺序进行排序，取排名在前的第三预设数量的文字作品，作为相似文字作品。需要说明的是，此处筛选的目的在于从已认证音频文字作品中筛选得到包含待认证音频文字作品中的全部或大部分切词的文字作品，以缩小认证比对的范围，节省服务器资源，进一步提高版权认证的效率。

然后，计算相似文字音频作品与待认证音频文字作品之间的第二最长公共子序列，其中，第二最长公共子序列中的“第二”并无实质含义，仅用于与上述第一最长公告子序列进行区分。进而根据第二最长公共子序列的长度计算相似文字音频作品与待认证数字作品之间的长度比值，得到第三计算结果，并检测第三计算结果中是否存在大于第三预设阈值

的长度比值，其中，第三预设阈值可以与第一预设阈值和第二预设阈值相同或不同，可根据实际需要设定，例如可设为 0.8，此处不作具体限定。若存在大于第三预设阈值的长度比值，则认为存在引用/抄袭关系，此时判定版权认证不通过；若不存在大于第三预设阈值的长度比值，则判定版权认证通过。

需要说明的是，在具体实施时，“计算相似音频文字作品与待认证数字作品之间的第二最长公共子序列，并根据第二最长公共子序列的长度计算相似音频文字作品与待认证数字作品之间的长度比值”与“检测是否存在大于第三预设阈值的长度比值”的过程可同时进行，即，依次计算各相似音频文字作品与待认证数字作品之间的第二最长公共子序列，然后根据第二最长公共子序列的长度计算相似音频文字作品与待认证数字作品之间的长度比值，进而该长度比值是否大于第三预设阈值，一旦检测到某一相似音频文字作品对应的长度比值大于第三预设阈值，则认为存在引用/抄袭关系，即可判定版权认证不通过，此时则无需计算其他相似音频文字作品与待认证数字作品之间的第二最长公共子序列及后续步骤，可节省服务器资源，进一步提高版权认证的效率。

进一步地，基于上述第一、第二和第四实施例，提出本申请版权认证方法的第五实施例。

在本实施例中，在步骤 S40 之后，该版权认证方法还可以包括：

步骤 A，在版权认证通过时，获取所述待认证数字作品的作品信息；

在本实施例中，在版权认证通过时，获取待认证数字作品的作品信息，其中，作品信息中的其中一部分信息（如作品作者、作品类型等）可以根据数字作品版权认证请求获取得到，也可以生成对应的提示窗口，在用户基于提示窗口填写对应的作品信息后获取得到；另一部分信息可由版权认证系统生成，例如认证时间、作品的 md5（Message-Digest Algorithm，信息摘要算法）值，其中认证时间可直接获取版权认证通过时的时间，作品的 md5 值可在获取到待认证数字作品后，通过对应的程序获取，例如，获取文件的 byte（字节）信息，第二步通过 MessageDigest（信息摘要）类进行 md5 加密，第三步转换成 16 进制的 md5 码值。作品信息可以包括但不限于：作品作者、作品类型、认证时间及待认证数字作品的 md5 值。

步骤 B，基于所述作品信息生成对应的版权认证信息，并根据所述版权认证信息生成数据上链请求；

然后，基于作品信息生成对应的版权认证信息，具体的，可以将作品信息转换成预设数据格式，如 json（一种轻量级的数据交换格式）数据结构，以得到版权认证信息。在生成版权认证信息之后，进而根据版权认证信息生成数据上链请求，具体的，可以采用 sha256（Secure Hash Algorithm 256，安全散列算法 256）算法生成该版权认证信息的哈希值，进而基于该哈希值生成数据上链请求。

步骤 C，将所述数据上链请求发送至版权认证联盟链，以使得所述版权认证联盟链基于所述数据上链请求完成对所述待认证数字作品的上链操作。

最后，将数据上链请求发送至版权认证联盟链，以使得版权认证联盟链基于数据上链请求完成对待认证数字作品的上链操作。其中，版权认证联盟链主要由版权认证机构节点、公证机构节点、司法机构节点以及外联节点组成，可以将数据上链请求发送至版权认证联盟链中的外联节点，然后通过所有节点一起基于通用的共识算法完成对待认证数字作品的上链操作，即将数据上链请求中的哈希值写入联盟链中永久保留。

当然，可以理解的是，在版权认证通过时，除上报版权认证联盟链对数据上链外，还可以将该版权认证通过的待认证数字作品存储至预设已认证作品库中，用于检测后续作品是否存在抄袭已认证作品的情况；同时，向用户端返回版权认证成功消息提示，已告知用户。

在本实施例中，版权认证系统在判定版权认证通过时，可生成对应的数据上链请求，并发送至版权认证联盟链，以使得版权认证联盟链基于数据上链请求完成上链操作，基于区块链具有不可修改的特性，可实现对数字作品版权的保护，后续一旦出现未经授权对他人作品进行传播的情况，版权所有者即可根据区块链上的认证信息对侵权行为进行申诉，降低维权的难度。

在现有技术中，也存在基于区块链的版权认证方案，例如，版权认证机构、公证机构、司法机构以及若干自媒体人分别作为一个节点，共同构成一个版权认证区块链平台，以对数字作品的版权进行审查认证。但是其对作品的审查以及版权认证环节还是在线下通过人工完成，而仅仅把最终的认证结果写入区块链平台，这样的方式并没有缩短认证周期，提高版权认证效率。同时，当前基于区块链的方案，多采用公链方式组织区块链，以此来保证系统运行的稳定性，此外还要求自媒体人或者创作者申请并加入区块链平台，增加创作者的使用成本，且他人可知晓各节点的操作行为，无法有效保证数据隐私性。对此，本申请还提供一种版权认证系统。

参照图 3，图 3 为本申请版权认证系统的系统架构示意图。

在本实施例中，如图 3 所示，版权认证系统包括版权认证设备和版权认证联盟链；当然，还可以包括用户端。

其中，版权认证设备为如图 1 所示的版权认证设备；用于执行上述版权认证方法实施例中的各步骤，具体的功能和实现过程可参照上述实施例，此处不作赘述。

所述版权认证联盟链，用于接收所述版权认证设备发送的数据上链请求；

基于所述数据上链请求获取待上链数据信息，并基于共识算法完成对所述待上链数据信息的上链操作。

在本实施例中，版权认证联盟链，可以用于接收版权认证设备发送的数据上链请求，其中，版权认证联盟链主要由版权认证机构节点、公证机构节点、司法机构节点以及外联节点组成，可通过外联节点接收版权认证设备发送的数据上链请求。进而基于数据上链请求获取待上链数据信息，其中，该待上链数据信息可以为基于版权认证信息生成的哈希值，然后基于共识算法完成对该哈希值的上链操作，即将该哈希值写入联盟链中永久保留。

此外，为保证版权认证的权威性和安全性，通常只对来自版权认证设备的请求进行响应，对应的，外联节点在接收到数据上链请求时，可获取对应的设备 IP (Internet Protocol , 互联网协议地址)，进而检测该设备 IP 是否为预设的版权认证设备的 IP，若是，则基于数据上链请求获取待上链数据信息，并基于共识算法完成对待上链数据信息的上链操作。

通过构建上述版权认证系统，可通过版权认证设备实现对数字作品版权的在线认证，可以针对不同类型的数字作品在秒级的时间复杂度内完成版权认证，相比于现有技术中通过人工进行版权认证，本申请可降低人力成本、缩短版权认证周期、提高版权认证效率，从而可及时对作品著作权人的版权进行保护。此外，本实施例中，自媒体人或者创作者无需申请并加入区块链平台、即可实现版权认证和保护，从而可减少创作者的使用成本，同时，他人也无法获取创作者的行为操作，可保证数据隐私性。

本申请还提供一种版权认证装置。

参照图 4，图 4 为本申请版权认证装置第一实施例的功能模块示意图。

如图 4 所示，所述版权认证装置包括：

第一获取模块 10，用于在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型；

处理模块 20，用于根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象；

审查模块 30，用于将所述目标输入对象输入至所述目标审查分类模型中，得到审查

结果，并基于所述审查结果判断审查是否通过；

版权认证模块 40，用于在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。

进一步地，所述处理模块 20 具体用于：

5 若所述作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型；

基于所述第一处理策略对所述待认证数字作品进行切词处理，得到第一切词文本；

将所述第一切词文本输入预设词向量模型，得到所述第一切词文本中各切词的第一词向量；

10 根据所述第一词向量得到所述待认证数字作品对应的第一文档向量，其中，所述目标输入对象为所述第一文档向量。

进一步地，所述处理模块 20 还具体用于：

若所述作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型；

15 基于所述第二处理策略对所述待认证数字作品进行预处理，得到输入图片，其中，所述预处理包括缩放处理和灰度处理，所述目标输入对象为所述输入图片。

进一步地，所述处理模块 20 还具体用于：

若所述作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型；

20 基于所述第三处理策略将所述待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品；

对所述转换后的待认证数字作品进行切词处理，得到第二切词文本；

将所述第二切词文本输入预设词向量模型，得到所述第二切词文本中各切词的第二词向量；

25 根据所述第二词向量得到所述转换后的待认证数字作品对应的第二文档向量，其中，所述目标输入对象为所述第二文档向量。

进一步地，所述目标审查分类模型包括多个，所述审查结果的数量与所述目标审查分类模型的数量相同，所述审查模块 30 具体用于：

检测多个所述审查结果是否均为审查合格；

30 若多个所述审查结果均为审查合格，则判定审查通过；

若多个所述审查结果中至少存在一个为审查不合格，则判定审查不通过。

进一步地，若所述作品类型为文字作品，所述版权认证模块 40 具体用于：

通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值；

35 根据所述第一相似度值从所述已认证文字作品中筛选得到第一预设数量的相似文字作品；

计算所述相似文字作品与所述待认证数字作品之间的第一最长公共子序列，并根据所述第一最长公共子序列的长度计算所述相似文字作品与所述待认证数字作品之间的长度比值，得到第一计算结果；

40 检测所述第一计算结果中是否存在大于第一预设阈值的长度比值；

若存在大于第一预设阈值的长度比值，则判定版权认证不通过；

若不存在大于第一预设阈值的长度比值，则判定版权认证通过。

进一步地，所述版权认证模块 40 还具体用于：

通过预设文档搜索引擎对所述待认证数字作品进行分词处理，得到分词集；

45 通过所述预设文档搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引，并

根据所述倒排索引结果计算所述分词集中各分词对应的分值；

对各分词的分值进行加和处理，得到所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。

进一步地，若所述作品类型为图片作品，所述版权认证模块 40 还具体用于：

5 通过预设图像检索引擎计算所述待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值；

根据所述第二相似度值从所述已认证图片作品中筛选得到第二预设数量的相似图片作品；

10 提取所述待认证数字作品的第一尺度不变特征变换 SIFT 特征向量，并提取所述相似图片作品的第二 SIFT 特征向量；

计算所述第一 SIFT 特征向量与所述第二 SIFT 特征向量之间的余弦距离，得到第二计算结果；

检测所述第二计算结果中是否存在大于第二预设阈值的余弦距离；

若存在大于第二预设阈值的余弦距离，则判定版权认证不通过；

15 若不存在大于第二预设阈值的余弦距离，则判定版权认证通过。

进一步地，若所述作品类型为音频作品，所述版权认证模块 40 还具体用于：

将所述待认证数字作品转换成文字作品类型，得到待认证音频文字作品；

通过预设文档检索引擎计算所述待认证音频文字作品与预设认证作品库中的已认证音频文字作品之间的第三相似度值；

20 根据所述第三相似度值从所述已认证音频文字作品中检索得到第三预设数量的相似音频文字作品；

计算所述相似音频文件作品与所述待认证音频文字作品之间的第二最长公共子序列，并根据所述第二最长公共子序列的长度计算所述相似音频文字作品与所述待认证音频文字作品之间的长度比值，得到第三计算结果；

25 检测所述第三计算结果中是否存在大于第三预设阈值的长度比值；

若存在大于第三预设阈值的长度比值，则判定版权认证不通过；

若不存在大于第三预设阈值的长度比值，则判定版权认证通过。

进一步地，所述版权认证装置还包括：

第二获取模块，用于在版权认证通过时，获取所述待认证数字作品的作品信息；

30 生成模块，用于基于所述作品信息生成对应的版权认证信息，并根据所述版权认证信息生成数据上链请求；

发送模块，用于将所述数据上链请求发送至版权认证联盟链，以使得所述版权认证联盟链基于所述数据上链请求完成对所述待认证数字作品的上链操作。

35 其中，上述版权认证装置中各个模块的功能实现与上述版权认证方法实施例中各步骤相对应，其功能和实现过程在此处不再一一赘述。

本申请还提供一种计算机可读存储介质，该计算机可读存储介质上存储有版权认证程序，所述版权认证程序被处理器执行时实现如以上任一项实施例所述的版权认证方法的步骤。

40 本申请计算机可读存储介质的具体实施例与上述版权认证方法各实施例基本相同，在此不作赘述。

45 需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者系统不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者系统所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包

括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

5 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在如上所述的一个存储介质(如 ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，空调器，或者网络设备等)执行本申请各个实施例所述的方法。

10 以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权 利 要 求 书

1、一种版权认证方法，其中，所述版权认证方法包括：

在接收到数字作品版权认证请求时，根据所述数字作品版权认证请求获取待认证数字作品和作品类型；

根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象；

将所述目标输入对象输入至所述目标审查分类模型中，得到审查结果，并基于所述审查结果判断审查是否通过；

在审查通过时，将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证。

2、如权利要求 1 所述的版权认证方法，其中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

若所述作品类型为文字作品，则确定对应的处理策略为第一处理策略，并确定目标审查分类模型为第一审查分类模型；

基于所述第一处理策略对所述待认证数字作品进行切词处理，得到第一切词文本；

将所述第一切词文本输入预设词向量模型，得到所述第一切词文本中各切词的第一词向量；

根据所述第一词向量得到所述待认证数字作品对应的第一文档向量，其中，所述目标输入对象为所述第一文档向量。

3、如权利要求 1 所述的版权认证方法，其中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

若所述作品类型为图片作品，则确定对应的处理策略为第二处理策略，并确定目标审查分类模型为第二审查分类模型；

基于所述第二处理策略对所述待认证数字作品进行预处理，得到输入图片，其中，所述预处理包括缩放处理和灰度处理，所述目标输入对象为所述输入图片。

4、如权利要求 1 所述的版权认证方法，其中，所述根据所述作品类型确定对应的处理策略和目标审查分类模型，基于所述处理策略对所述待认证数字作品进行处理，得到目标输入对象的步骤包括：

若所述作品类型为音频作品，则确定对应的处理策略为第三处理策略，并确定目标审查分类模型为第三审查分类模型；

基于所述第三处理策略将所述待认证数字作品转换成文字作品类型，得到转换后的待认证数字作品；

对所述转换后的待认证数字作品进行切词处理，得到第二切词文本；

将所述第二切词文本输入预设词向量模型，得到所述第二切词文本中各切词的第二词向量；

根据所述第二词向量得到所述转换后的待认证数字作品对应的第二文档向量，其中，所述目标输入对象为所述第二文档向量。

5、如权利要求 1 至 4 中任一项所述的版权认证方法，其中，所述目标审查分类模型包括多个，所述审查结果的数量与所述目标审查分类模型的数量相同，所述基于所述审查结果判断审查是否通过的步骤包括：

检测多个所述审查结果是否均为审查合格；

若多个所述审查结果均为审查合格，则判定审查通过；

若多个所述审查结果中至少存在一个为审查不合格，则判定审查不通过。

6、如权利要求 1 所述的版权认证方法，其中，若所述作品类型为文字作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值；

根据所述第一相似度值从所述已认证文字作品中筛选得到第一预设数量的相似文字作品；

计算所述相似文字作品与所述待认证数字作品之间的第一最长公共子序列，并根据所述第一最长公共子序列的长度计算所述相似文字作品与所述待认证数字作品之间的长度比值，得到第一计算结果；

检测所述第一计算结果中是否存在大于第一预设阈值的长度比值；

若存在大于第一预设阈值的长度比值，则判定版权认证不通过；

若不存在大于第一预设阈值的长度比值，则判定版权认证通过。

7、如权利要求 6 所述的版权认证方法，其中，所述通过预设文档搜索引擎计算所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值的步骤包括：

通过预设文档搜索引擎对所述待认证数字作品进行分词处理，得到分词集；

通过所述预设文档搜索引擎对预设认证作品库中的已认证文字作品进行倒排索引，并根据所述倒排索引结果计算所述分词集中各分词对应的分值；

对各分词的分值进行加和处理，得到所述待认证数字作品与预设认证作品库中的已认证文字作品之间的第一相似度值。

8、如权利要求 1 所述的版权认证方法，其中，若所述作品类型为图片作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

通过预设图像检索引擎计算所述待认证数字作品与预设认证作品库中的已认证图片作品之间的第二相似度值；

根据所述第二相似度值从所述已认证图片作品中筛选得到第二预设数量的相似图片作品；

提取所述待认证数字作品的第一尺度不变特征变换 SIFT 特征向量，并提取所述相似图片作品的第二 SIFT 特征向量；

计算所述第一 SIFT 特征向量与所述第二 SIFT 特征向量之间的余弦距离，得到第二计算结果；

检测所述第二计算结果中是否存在大于第二预设阈值的余弦距离；

若存在大于第二预设阈值的余弦距离，则判定版权认证不通过；

若不存在大于第二预设阈值的余弦距离，则判定版权认证通过。

9、如权利要求 1 所述的版权认证方法，其中，若所述作品类型为音频作品，所述将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对，以进行版权认证的步骤包括：

将所述待认证数字作品转换成文字作品类型，得到待认证音频文字作品；

通过预设文档搜索引擎计算所述待认证音频文字作品与预设认证作品库中的已认证音频文字作品之间的第三相似度值；

根据所述第三相似度值从所述已认证音频文字作品中检索得到第三预设数量的相似音频文字作品；

计算所述相似音频文件作品与所述待认证音频文字作品之间的第二最长公共子序

列,并根据所述第二最长公共子序列的长度计算所述相似音频文字作品与所述待认证音频文字作品之间的长度比值,得到第三计算结果;

检测所述第三计算结果中是否存在大于第三预设阈值的长度比值;

若存在大于第三预设阈值的长度比值,则判定版权认证不通过;

5 若不存在大于第三预设阈值的长度比值,则判定版权认证通过。

10、如权利要求1至4、6至9中任一项所述的版权认证方法,其中,所述版权认证方法还包括:

在版权认证通过时,获取所述待认证数字作品的作品信息;

10 基于所述作品信息生成对应的版权认证信息,并根据所述版权认证信息生成数据上链请求;

将所述数据上链请求发送至版权认证联盟链,以使得所述版权认证联盟链基于所述数据上链请求完成对所述待认证数字作品的上链操作。

11、一种版权认证装置,其中,所述版权认证装置包括:

15 第一获取模块,用于在接收到数字作品版权认证请求时,根据所述数字作品版权认证请求获取待认证数字作品和作品类型;

处理模块,用于根据所述作品类型确定对应的处理策略和目标审查分类模型,基于所述处理策略对所述待认证数字作品进行处理,得到目标输入对象;

审查模块,用于将所述目标输入对象输入至所述目标审查分类模型中,得到审查结果,并基于所述审查结果判断审查是否通过;

20 版权认证模块,用于在审查通过时,将所述待认证数字作品与预设认证作品库中的已认证数字作品进行比对,以进行版权认证。

12、一种版权认证设备,其中,所述版权认证设备包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的版权认证程序,所述版权认证程序被所述处理器执行时实现如权利要求1至10中任一项所述的版权认证方法的步骤。

25 13、一种版权认证系统,其中,所述版权认证系统包括版权认证设备和版权认证联盟链;其中,

所述版权认证设备为如权利要求12所述的版权认证设备;

所述版权认证联盟链,用于接收所述版权认证设备发送的数据上链请求;

30 基于所述数据上链请求获取待上链数据信息,并基于共识算法完成对所述待上链数据信息的上链操作。

14、一种计算机可读存储介质,其中,所述计算机可读存储介质上存储有版权认证程序,所述版权认证程序被处理器执行时实现如权利要求1至10中任一项所述的版权认证方法的步骤。

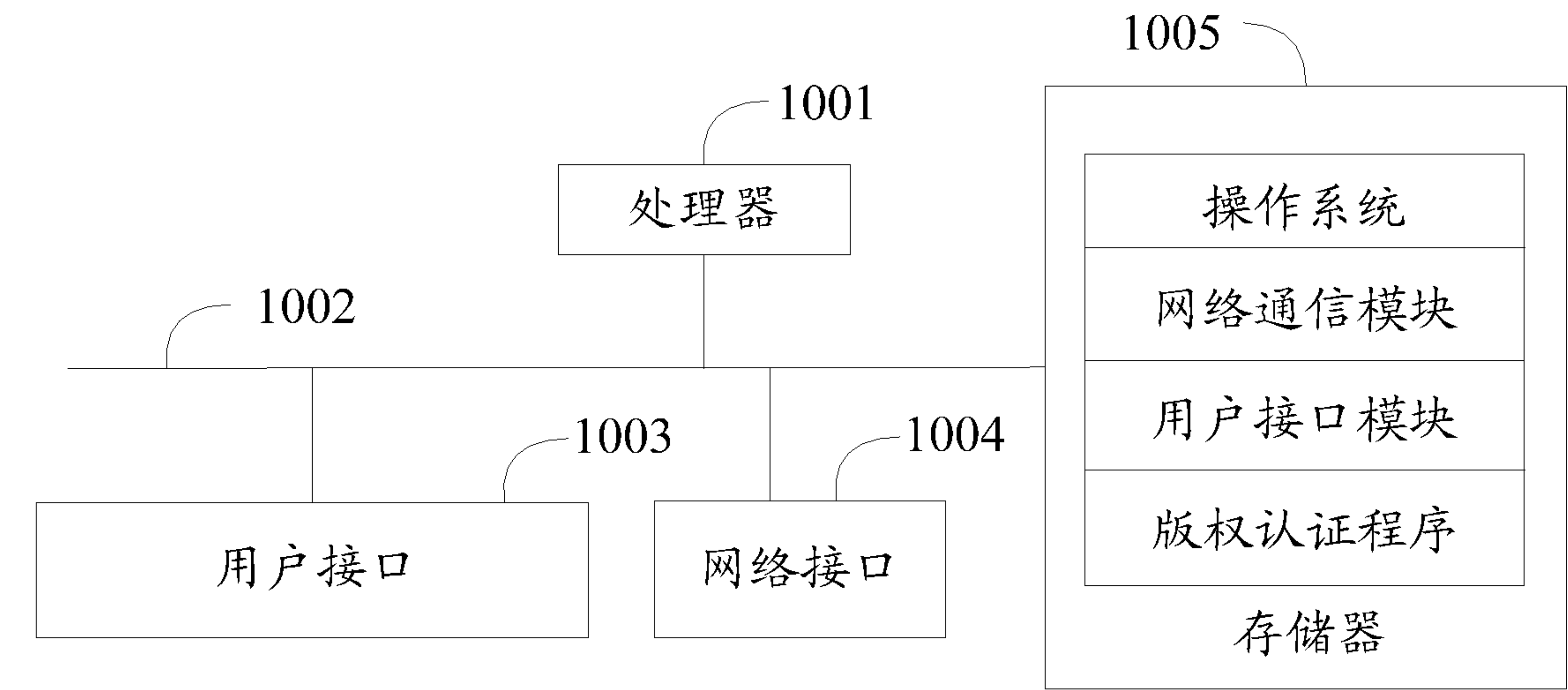


图 1

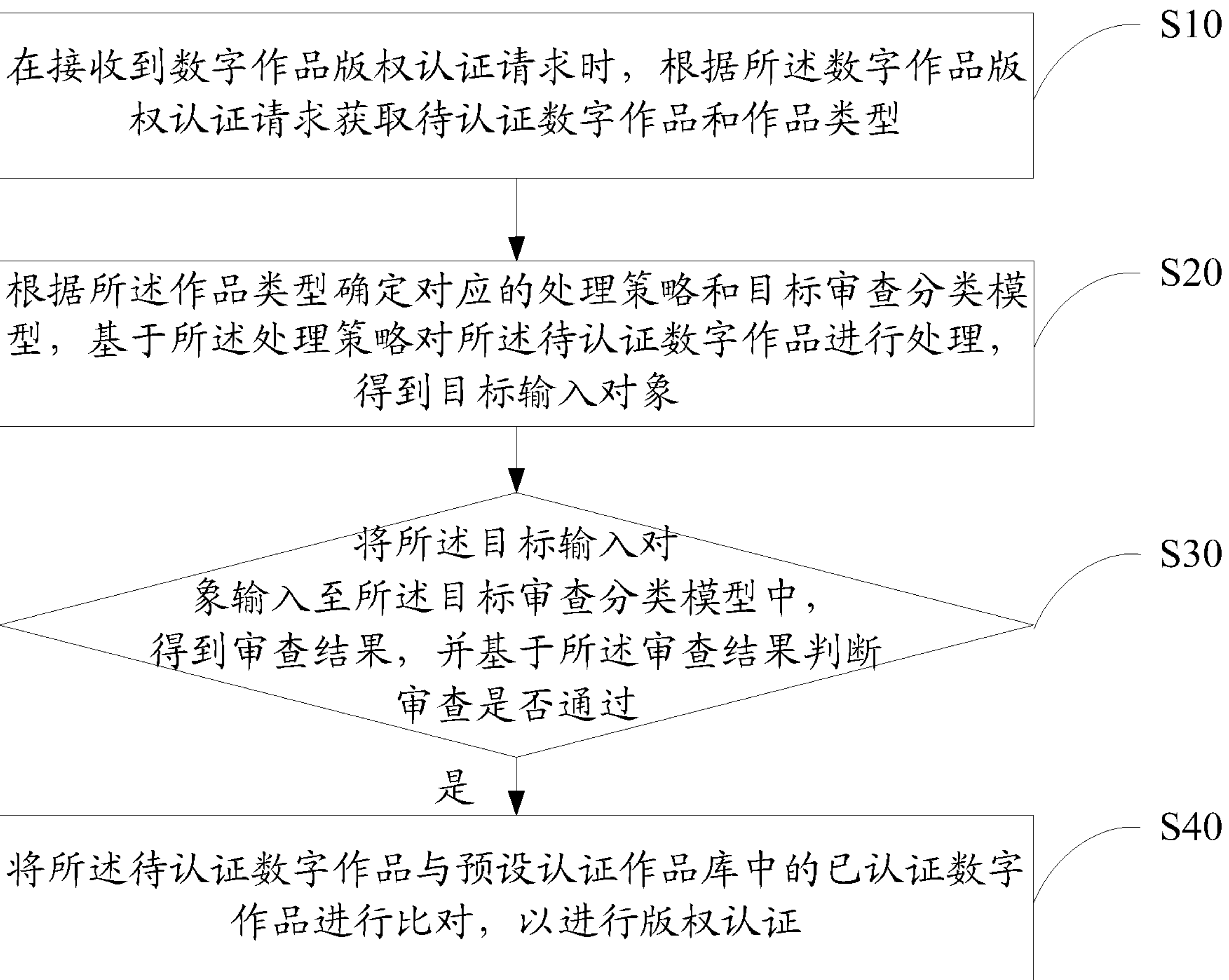


图 2

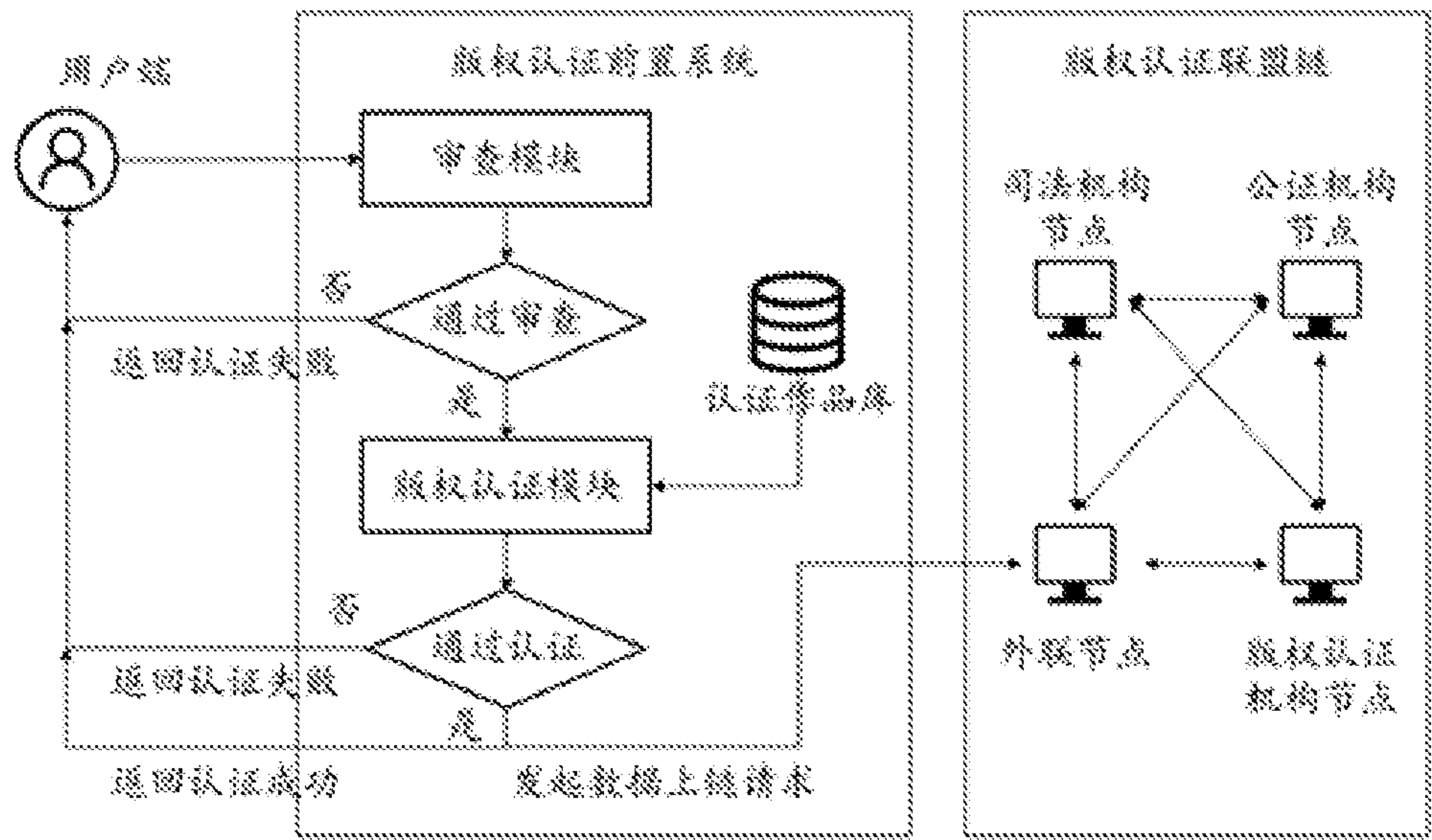


图 3



图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/126232

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/10(2013.01)i; G06K 9/62(2006.01)n

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)

G06F G06K

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)

CNPAT, CNKI, WPI, EPODOC: 版权, 认证, 抄袭, 检测, 公共, 序列, 长度, 比值, 距离, 图片, 相似, 上传, 类型, 不良, 敏感, 过滤, 文字, 文档, copyright, authentic+, common, sequence, length, ratio, distance, image, upload, similar+, type, harmful, sensitive, filter, literal, sift

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101645117 A (WUHAN UNIVERSITY) 10 February 2010 (2010-02-10) description, page 4, last paragraph to page 7, last paragraph	1-14
Y	CN 106874253 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 20 June 2017 (2017-06-20) description, paragraphs 57-104	1-14
Y	CN 107832384 A (BEIJING ANNE FULL COPYRIGHT TECHNOLOGY DEVELOPMENT CO., LTD.) 23 March 2018 (2018-03-23) description, paragraphs 24-56	2-4,12-14
Y	CN 101826099 A (BLUEDON INFORMATION SECURITY TECHNOLOGY CO., LTD.) 08 September 2010 (2010-09-08) description, paragraphs 35-53	6-7,9,12-14
Y	CN 105550381 A (BEIJING UNIVERSITY OF TECHNOLOGY) 04 May 2016 (2016-05-04) description, paragraphs 30-89	8,12-14
A	CN 109145529 A (CHONGQING INDUSTRY POLYTECHNIC COLLEGE) 04 January 2019 (2019-01-04) entire document	1-14

☒ 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

14 January 2021

Date of mailing of the international search report

04 February 2021

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

International application No.

PCT/CN2020/126232

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110781460 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 11 February 2020 (2020-02-11) claims 1-13	1-14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2020/126232

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	101645117	A	10 February 2010	CN	101645117	B	30 November 2011
CN	106874253	A	20 June 2017	None			
CN	107832384	A	23 March 2018	None			
CN	101826099	A	08 September 2010	CN	101826099	B	05 September 2012
CN	105550381	A	04 May 2016	CN	105550381	B	05 April 2019
CN	109145529	A	04 January 2019	None			
CN	110781460	A	11 February 2020	None			

国际检索报告		国际申请号 PCT/CN2020/126232
A. 主题的分类 G06F 21/10(2013.01)i; G06K 9/62(2006.01)n 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. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06F G06K 包含在检索领域中的除最低限度文献以外的检索文献 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)) CNPAT, CNKI, WPI, EPDOC: 版权, 认证, 抄袭, 检测, 公共, 序列, 长度, 比值, 距离, 图片, 相似, 上传, 类型, 不良, 敏感, 过滤, 文字, 文档, copyright, authentic+, common, sequence, length, ratio, distance, image, upload, similar+, type, harmful, sensitive, filter, literal, sift		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	CN 101645117 A (武汉大学) 2010年 2月 10日 (2010 - 02 - 10) 说明书第4页倒数第1段-第7页最后一段	1-14
Y	CN 106874253 A (腾讯科技深圳有限公司) 2017年 6月 20日 (2017 - 06 - 20) 说明书第57-104段	1-14
Y	CN 107832384 A (北京安妮全版权科技发展有限公司) 2018年 3月 23日 (2018 - 03 - 23) 说明书第24-56段	2-4, 12-14
Y	CN 101826099 A (蓝盾信息安全技术股份有限公司) 2010年 9月 8日 (2010 - 09 - 08) 说明书第35-53段	6-7, 9, 12-14
Y	CN 105550381 A (北京工业大学) 2016年 5月 4日 (2016 - 05 - 04) 说明书第30-89段	8, 12-14
A	CN 109145529 A (重庆工业职业技术学院) 2019年 1月 4日 (2019 - 01 - 04) 全文	1-14
PX	CN 110781460 A (深圳前海微众银行股份有限公司) 2020年 2月 11日 (2020 - 02 - 11) 权利要求1-13	1-14
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “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) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “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, 但为了解发明之理论或原理的在后文件 “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 “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 “&” 同族专利的文件		
国际检索实际完成的日期 2021年 1月 14日		国际检索报告邮寄日期 2021年 2月 4日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 巩瑜 电话号码 86-(10)-53961382

国际申请号
PCT/CN2020/126232

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)	
CN	101645117	A	2010年 2月 10日	CN	101645117 B	2011年 11月 30日
CN	106874253	A	2017年 6月 20日	无		
CN	107832384	A	2018年 3月 23日	无		
CN	101826099	A	2010年 9月 8日	CN	101826099 B	2012年 9月 5日
CN	105550381	A	2016年 5月 4日	CN	105550381 B	2019年 4月 5日
CN	109145529	A	2019年 1月 4日	无		
CN	110781460	A	2020年 2月 11日	无		