(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
2019 年 10 月 17 日 (17.10.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/196228 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2018/095549

(22) 国际申请日: 2018 年 7 月 13 日 (13.07.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810313040.X 2018年4月9日 (09.04.2018) 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 518033 (CN)。

(72) 发明人: 韩梅 (HAN, Mei); 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 518033 (CN)。 张安元 (ZHANG, Anyuan); 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 518033 (CN)。 邓华威 (DENG, Huawei); 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 518033 (CN)。 王科 (WANG, Ke); 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 518033 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路6号4501房（部位：自编01-03和08-12单元）（仅限办公用途）, Guangdong 510623 (CN)。

(54) **Title:** SYSTEM INFORMATION PROCESSING METHOD, APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) **发明名称:** 制度信息处理方法、装置、计算机设备和存储介质

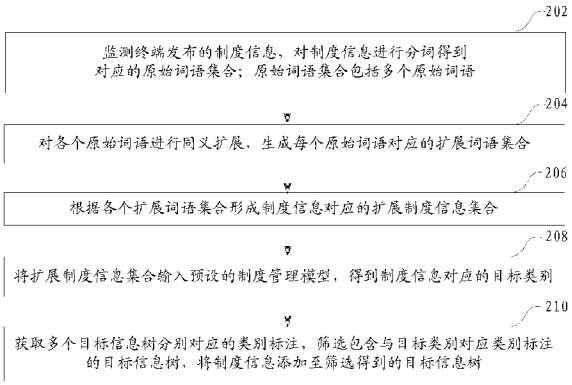


图 2

202 Monitor system information issued by a terminal, and segment the system information to obtain a corresponding original word set, the original word set comprising a plurality of original words;

204 Synonymously extend each original word and generate an extende word set corresponding to each original word

206 Form an extended system information set corresponding to the system information according to each extended word set

208 Input the extended system information set into a preset system management model, and obtain a target category corresponding to the system information

210 Acquire category labels respectively corresponding to a plurality of target information trees, filter a target information tree comprising a category label that corresponds to the target category, and add the system information to the target information tree that is obtained by filtering

(57) **Abstract:** A system information processing method, comprising: monitoring system information issued by a terminal, and segmenting the system information to obtain a corresponding original word set, the original word set comprising a plurality of original words; synonymously extending each original word and generating an extended word set corresponding to each original word; forming an extended system information set corresponding to the system information according to each extended word set; inputting the extended system information set into a preset system management model, and obtaining a target category corresponding to the system information; acquiring category labels respectively corresponding to a plurality of target information trees, filtering a target information tree comprising a category label that corresponds to the target category, and adding the system information to the target information tree that is obtained by filtering.

(57) **摘要**: 一种制度信息处理方法，包括：监测终端发布的制度信息，对制度信息进行分词得到对应的原始词语集合；原始词语集合包括多个原始词语；对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；根据各个扩展词语集合形成制度信息对应的扩展制度信息集合；将扩展制度信息集合输入预设的制度管理模型，得到制度信息对应的目标类别；获取多个目标信息树分别对应的类别标注，筛选包含与目标类别对应类别标注的目标信息树，将制度信息添加至筛选得到的目标信息树。

WO 2019/196228 A1

[见续页]

(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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

制度信息处理方法、装置、计算机设备和存储介质

本申请要求于2018年4月9日提交中国专利局，申请号为201810313040X，申请名称为“制度信息处理方法、装置、计算机设备和存储介质”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

本申请涉及一种制度信息处理方法、装置、计算机设备和存储介质。

背景技术

企业标准化是对企业生产经营与管理等活动中的重复性事物和概念，通过制订、发布和实施制度规范达到统一，以提高企业管理水平。制度规范（以下简称“制度”）是员工在生产经营活动中须共同遵守的规定和准则，包括法律与政策、企业组织结构、管理制度、岗位职责、技术标准、工作流程等规范文件。为了满足各个岗位的工作需求，企业需要从不同维度对制度进行分类管理，构建多个不同的信息树，如技术标准信息树、法律政策信息树等，以使不同类和用途的制度形成不同的制度体系。当有新制度发布时，需要将新发布的制度纳入相应的信息树。同一种制度可能同时隶属于多个不同的信息树。随着企业规模增大，相应的制度信息和信息树均越来越多。但发明人意识到，在传统的方式中，通过人工对海量制度信息进行分类管理，不仅效率低，且易出错。

发明内容

根据本申请公开的各种实施例，提供一种制度信息处理方法、装置、计算机设备和存储介质。

一种制度信息处理方法包括：监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语；对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合；将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别；及获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标

注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

一种制度信息处理装置包括:

信息扩展模块，用于监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语；对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合;

信息分类模块，用于将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别；及

信息归档模块，用于获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

一种计算机设备，包括存储器和一个或多个处理器，存储器中存储有计算机可读指令，计算机可读指令被处理器执行时实现本申请任意一个实施例中提供的制度信息处理方法的步骤。

一个或多个存储有计算机可读指令的非易失性存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现本申请任意一个实施例中提供的制度信息处理方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

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，下面描述中的附图仅仅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它的附图。

图 1 为根据一个或多个实施例中制度信息处理方法的应用场景图。

图 2 为根据一个或多个实施例中制度信息处理方法的流程示意图。

图 3 为根据一个或多个实施例中制度信息处理方法中目标信息树的示意图。

图 4 为根据一个或多个实施例中构建关联信息树步骤的流程示意图。

图 5 为根据一个或多个实施例中制度信息处理方法中关联信息树的示意图。

图 6 为根据一个或多个实施例中制度信息处理装置的结构框图。

图 7 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处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。

本申请提供的制度信息处理方法，可以应用于如图 1 所示的应用环境中。终端 102 与服务器 104 通过网络进行通信。终端 102 可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

服务器 104 中存储了多种目标信息树。每种目标信息树具有对应的类别标注。服务器 104 对终端 102 是否发布新的制度信息进行监测，当监测到终端 102 发布了新的制度信息时，服务器 104 对制度信息进行分类，将制度信息纳入相应的目标信息树。具体的，服务器 104 对制度信息进行分词，得到包括多个原始词语的原始词语集合。服务器 104 获取各个原始词语对应的同义词，将原始词语与对应的同义词形成扩展词语集合。每个原始词语都存在对应的扩展词语集合。服务器 104 按与制度信息中各个原始词语出现的顺序，从各个原始词语对应的扩展词语集合中任意选择一个词语，按顺序形成一个扩展制度信息。当从扩展词语集合中选择不同的词语时，则形成不同的扩展制度信息，不同的扩展制度信息组成扩展制度信息集合。服务器 104 将扩展制度信息集合输入已训练的制度管理模型，利用制度管理模型确定制度信息对应的目标类别。服务器 104 获取与目标类别对应的类别标注，筛选包含获取到的类别标注的信息节点，将制度信息添加至筛选得到的目标信息树。对制度信息进行分词和同义扩展，提高了制度信息的有效覆盖范围，从而在后续输入制度管理模型后，可提高目标类别的精准性，进而可准确将制度信息纳入相应目标信息树，提高制度信息分类效率和准确率。

在其中一个实施例中，如图 2 所示，提供了一种制度信息处理方法，以该方法应用于图 1 中的服务器为例进行说明，包括以下步骤：

步骤 202，监测终端发布的制度信息，对制度信息进行分词得到对应的原始

词语集合；原始词语集合包括多个原始词语。

服务器对第一终端是否发布新的制度信息进行监测。制度信息包括制度描述信息及关联的制度文件。制度描述信息包括制度编码、制度名称、制度级别、发布单位、发布日期、适用对象标识或信息摘要等。制度信息可以是文本信息，也可以是语音信息、图像信息、视频信息等。如果是语音信息、图像信息或视频信息，则可先通过语音识别或图像处理，将语音信息、图像信息和视频信息转化为文本信息。制度文件包括多项制度条款以及每项制度条款对应的适用对象标识。适用对象标识是指需要执行或了解该制度的对象的标识信息，可以是岗位标识或机构标识等。

当监测到第一终端发布了新的制度信息时，服务器对制度信息进行分类。具体的，服务器通过分词算法对制度信息进行分词，得到原始词语集合。原始词语集合包括多个原始词语。在其中一个实施例中，得到各个原始词语后，去除停用词、语气词、标点符号等对分类影响作用小的词语，从而提高后续特征提取的效率。停用词指的是制度信息中出现频率超过预设阈值但实际意义不大的词，如我，的，他等。

在其中一个实施例中，在对制度信息进行分词得到对应的原始词语集合之前，还包括：检测制度描述信息是否包含类别信息；若包含，则根据类别信息将制度信息添加至相应的目标信息树；否则，对制度信息进行分词得到对应的原始词语集合。

终端在发布制度信息时，也可以预先标明制度信息的类别信息，以便服务器可以根据该类别信息，将制度信息纳入相应的目标信息树。若制度描述信息并未包含制度信息的类别信息，则可以按照本申请提供的制度信息处理方法对制度信息进行分类管理。

步骤204，对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合。

服务器分别获取原始词语集合中各个原始词语对应的同义词，将原始词语与对应的同义词形成扩展词语集合。每个原始词语都存在对应的扩展词语集合。同义词是指与原始词语含义相同或相近的词语，如原始词语为“不得”，同义词可为“切勿”、“禁止”、“避免”、“杜绝”等，将原始词语与对应的同义词形成扩展词语集合，如原始词语“不得”对应的扩展词语集合为{不得，切勿，禁止，

避免，杜绝}。如原始词语集合为{a,b,c}，则原始词语集合中的每个原始词语都存在对应的扩展词语集合，如 a 对应的扩展词语集合为{a,a1,a2 }，b 对应的扩展词语集合为{b,b1,b2,b3}，c 对应的扩展词语集合为{c,c1,c2 }。

步骤 206，根据各个扩展词语集合形成制度信息对应的扩展制度信息集合。

服务器按照与制度信息中各个原始词语出现的顺序，从各个原始词语对应的扩展词语集合中任意选择一个词语，按顺序形成一个扩展制度信息。当从扩展词语集合中选择不同的词语时，则形成不同的扩展制度信息，不同的扩展制度信息组成扩展制度信息集合。

在其中一个实施例中，服务器对各个原始词语对应的扩展词语集合求笛卡尔积，形成由不同的扩展制度信息组成的扩展制度信息集合。两个集合 X 和 Y 的笛卡尔积，又称直积，表示为 X × Y。第一个对象是 X 的成员而第二个对象是 Y 的所有可能有序对的其中一个成员。

步骤 208，将扩展制度信息集合输入预设的制度管理模型，得到制度信息对应的目标类别。

制度管理模型用于根据输入从多个候选类型中确定与输入对应的目标类别。制度管理模型可以是通过逻辑回归算法、支持向量机算法等训练得到的模型。制度管理模型内部可以由多个子管理模型连接形成。由于已训练的制度管理模型的输入是经过扩展了的扩展制度信息集合，扩展后的各个扩展制度信息表达了与制度信息相同或相近的含义，提高了制度信息的有效覆盖范围，从而在后续输入已训练的制度管理模型后，可提高目标类别的精准性。

步骤 210，获取多个目标信息树分别对应的类别标注，筛选包含与目标类别对应类别标注的目标信息树，将制度信息添加至筛选得到的目标信息树。

服务器中存储了多种目标信息树。如图 3 所示，每种目标信息树包括多个信息节点及每个信息节点关联的制度文件。制度文件可以是多种格式的文件，如 pdf 文档、jpg 图像、xls 表格、mp3 音频或 avi 视频等。不同的信息节点在目标信息树中可以按照发布时间先后排列。容易理解，一项制度信息也可以不具有关联的制度文件，也还可以具有多个关联的制度文件，对此不作限制。

每种目标信息树具有对应的类别标注。类别标注用于标识相应目标信息树能够包含的信息节点的类别，如行政管理类、销售管理类或风险管理类等。服务器获取与目标类别对应的类别标注，筛选包含获取到的类别标注的一种或多

种目标信息树。服务器根据制度描述信息生成信息节点。例如，可以将制度编号和/或制度名称作为信息节点。服务器将制度文件关联至该信息节点，将关联有制度文件的信息节点添加至筛选得到的目标信息树。

在其中一个实施例中，制度信息包括制度描述信息和制度文件；将制度信息添加至筛选得到的目标信息树包括：根据制度描述信息生成信息节点；检测筛选得到的目标信息树中是否已存在相同的信息节点；若不存在，则将信息节点添加至相应的目标信息树，将制度文件关联至信息节点。

若筛选得到的关联信息树中已经存在相应的信息节点，则服务器只需将制度文件关联至已存在相应的信息节点。在另一个实施例中，服务器根据制度描述信息判断生成的信息节点与已存在的相同信息节点属于并列节点还是父子节点。当生成的信息节点与已存在的相同信息节点属于并列节点时，服务器对生成的信息节点与已存在的相同信息节点进行区别标记，将区别标记后的信息节点添加至相应的目标信息树，将制度文件关联至区别标记后的信息节点。

当生成的信息节点与已存在的相同信息节点属于并列节点时，服务器根据制度描述信息对生成的信息节点进行描述限定，即在制度描述信息中提取关键词，利用提取到的关键词对生成的信息节点进行语义扩充。例如，根据制度名称生成的信息节点为“公司福利管理制度”，在制度描述信息中提取关键词“研发部”，则语义扩充后的信息节点可以是“公司研发部福利管理制度”。服务器将语义扩充后的信息节点作为已存在的相同信息节点的子节点添加至相应的目标信息树，将制度文件关联至该子节点。

本实施例中，通过监测新发布的制度信息，对制度信息进行分词得到对应的原始词语集合；通过获取原始词语集合中各个原始词语对应的同义词，可以利用原始词语与对应的同义词形成扩展词语集合；根据各个原始词语对应的扩展词语集合，可以形成制度信息对应的扩展制度信息集合；将扩展制度信息集合输入已训练的制度管理模型得到制度信息对应的目标类别；将目标类别与预存储的多个目标信息树分别对应的类别标注进行匹配，可以筛选得到能够包含该制度信息的目标信息树，将制度信息添加至筛选得到的目标信息树。先形成每个原始词语对应的扩展词语集合，再通过扩展词语集合形成扩展制度信息集合，大大提高了扩展制度信息的扩展度，扩展后的各个扩展制度信息表达了与制度信息相同或相近的含义，提高了制度信息的有效覆盖范围，从而在后续输

入已训练的制度管理模型后，可提高目标类别的精准性，进而可以准确将制度信息纳入相应的目标信息树，提高制度信息分类效率和准确率。

在其中一个实施例中，制度管理模型的生成步骤包括：获取训练样本数据；训练样本数据包括多个样本制度信息及分别对应的类别标注；对各个样本制度信息进行分词和同义扩展处理，得到每个样本制度信息分别对应的扩展样本制度信息集合；根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，得到制度管理模型。

训练样本数据可以是已发布的多种样本制度信息。每种样本制度信息都有对应的类别标注，用于描述样本制度信息的实际类别。例如，样本制度信息对应的制度名称为“考勤注意事项”，则该样本制度信息对应的类别标注可以是“行政管理”。训练样本数据包括所有可能的类别对应的样本制度信息，以保证各个类别确定的准确性。在一个具体的实施例中，训练样本数据包括 476 个样本制度信息，类别标注总数为 57。

服务器通过分词算法对各个训练样本信息进行分词得到各个词语，各个词语组成各个训练样本信息对应的原始训练词语集合。服务器获取每个原始训练词语的同义词，将原始训练词语与对应的同义词形成扩展训练词语集合。扩展训练词语集合包括多组

服务器先获取其中一个训练样本信息作为当前训练样本信息，获取当前训练样本信息对应的各个原始训练词语，获取各个原始训练词语对应的扩展训练词语集合，然后按与当前训练样本信息中各个原始训练词语出现的顺序，从各个原始训练词语对应的扩展训练词语集合中任意选择一个词语，按顺序形成一个扩展样本制度信息。不同的扩展样本制度信息组成扩展样本制度信息集合。各个样本制度信息都有对应的扩展样本制度信息集合。在其中一个实施例中，服务器对各个原始训练词语对应的扩展训练词语集合求笛卡尔积，形成得到每个样本制度信息分别对应的扩展样本制度信息集。

支持向量机算法是一种用来进行模式识别，模式分类的机器学习算法。支持向量机的主要思想是：建立一个最优决策超平面，使得该平面两侧距离该平面最近的两类样本之间的距离最大化，从而对分类问题提供良好的泛化能力。对于一个多维的样本集，系统随机产生一个超平面并不断移动，对样本进行分类，直到训练样本中属于不同类别的样本点正好位于该超平面的两侧，满足该

条件的超平面可能有很多个，支持向量机算法在保证分类精度的同时，寻找到这样一个超平面，使得超平面两侧的空白区域最大化，从而实现对线性可分样本的最优分类，支持向量机算法是一种有监督的训练方法。在其中一个实施例中，制度管理模型由多个子管理模型连接形成。

本实施例中，对大量已发布制度信息进行分词和同义扩展处理，处理得到的扩展样本制度信息集合，大大提高了样本制度信息的有效覆盖范围；将扩展样本制度信息集合输入制度管理模型，并基于支持向量机算法对制度管理模型进行训练，可提高制度管理模型的分类精准性。

在其中一个实施例中，扩展样本制度信息集合包括多组扩展样本制度信息；根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练包括：获取特征项，计算特征项在一组扩展样本制度信息的词频权重；计算特征项在整个训练样本数据中的文档频率；根据词频权重和文档频率计算特征项对应的特征权重；根据特征权重选择特征项作为相应扩展样本制度信息的特征词；根据特征词提取各个扩展样本标准信息的特征。

特征项可以是一组扩展样本制度信息中的任一个词语。词频权重指的是特征项在该组扩展样本制度信息中出现的频率。可以理解的是，如果扩展样本制度信息中存在特征项的同义词，则也算出现。词频权重通常被归一化，可表示为TF_{ij}，其中 i 表示特征项对应的标识，j 表示类别标识。文档频率DF_i是一个词语普遍重要性的度量，可以由特征项所在的扩展样本制度信息数目除以训练样本数据中所有训练样本信息的总数目得到。

如果该特征项在扩展样本制度信息中出现的次数越多，表明该特征项对扩展样本制度信息的影响力度越大，即特征权重与词频权重成正比。如果该特征项出现的扩展样本制度信息的数量越多，表明，该特征项对信息分类的作用越小，即特征权重与文档频率成反比。在其中一个实施例中，特征权重$w_{t_i} = TF_{ij} * \log(\frac{N}{DF_i})$，其中 N 表示训练样本数据中所有训练样本信息的总数目。

如果特征权重超过预设阈值，则说明此特征项是这一组扩展样本制度信息的重要词语，可将此特征项作为此扩展样本制度信息的特征词。可根据确定的各个特征词提取扩展样本制度信息集合中各个扩展样本制度信息的特征。对于一个扩展样本制度信息，特征词可以包括一个或多个。

本实施例中，通过统计扩展样本制度信息中每个词语的词频权重和文档频

率，确定该词语代表扩展样本制度信息的特征权重，根据特征权重可以提取扩展样本制度信息集合中每个扩展样本制度信息的特征，从而使制度管理模型能够从基于多样化的语言描述的制度信息中准确提取其特征，进而进行准确分类。

在其中一个实施例中，制度信息包括制度描述信息及关联的制度文件；制度文件包括多个制度条款以及分别对应的适用对象标识；关联信息树具有对应的适用对象标识。该方法还包括构建关联信息树的步骤。如图 4 所示，构建关联信息树的步骤包括:

步骤 402，对制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件。

为了满足所有岗位的工作需求，企业可能将适用于不同岗位的制度信息全部记录至同一个制度文件中，使得用户只能基于制度文件全部信息内容进行制度查询，进而使制度信息查询效率降低。本实施例针对不同岗位构建不同的关联信息树。具体的，服务器根据制度文件中每个制度条款对应的适用对象标识，对制度文件中多个制度条款进行拆分，生成每个适用对象标识分别对应的制度子文件。例如，制度文件 A 包括 X1~X4 四项制度条款。其中，X1 对应适用对象标识包括甲和乙，X2 对应适用对象标识包括甲，X3 对应适用对象标识包括甲、乙、丙、丁和戊，X4 对应适用对象标识包括甲和丁。制度文件 A 共包括甲、乙、丙、丁和戊五个适用对象标识，对应的拆分得到五个制度子文件 A1~A5。其中，适用对象标识甲对应的制度子文件 A1 包括 X1~X4 四项制度条款；适用对象标识乙对应的制度子文件 A2 包括 X1 和 X3 两项制度条款；如此类推。

步骤 404，获取目标信息树对应的多个关联信息树。

每种目标信息树具有对应的多个关联信息树。目标信息树中每个信息节点具有对应的一个或多个适用对象标识。目标信息树中不同适用对象标识分别具有对应的一个关联信息树。换言之，目标信息树中包含适用对象标识的数量与对应的关联信息树的数量相等，从而每个适用对象标识对应岗位具有对应的关联信息树。

目标信息树用于记录适用于企业全部岗位的制度信息。而关联信息树则只需记录适用于一个岗位的制度信息。每种关联信息树具有对应的适用对象标识。如图 5 所示，岗位 1 无需执行或了解信息节点 4 和信息节点 9 对应的制度，则适用对象标识“岗位 1”对应的关联信息树，相对图 3 目标信息树不存在信息节

点 4 和信息节点 9。容易理解，关联信息树中多个信息节点的目录层级，并非一定与目标信息树一致，可以自适应调整。关联信息树仍存在的其他信息节点关联的制度文件记录的内容，与目标信息树中相应信息节点关联的制度文件记录的内容可以不同。

步骤 406，根据适用对象标识，将制度描述信息及制度子文件添加至相应的关联信息树。

服务器将制度信息添加至相应的目标信息树后，服务器根据制度文件记录的适用对象标识，获取目标信息树对应的相应关联信息树。容易理解，服务器只需获取制度文件记录的适用对象标识对应的关联信息树。例如，制度信息分类添加至三种目标信息树，其中包括目标信息树 M。目标信息树 M 对应适用对象标识包括甲、乙、丙、丁、戊和己，假设依上述举例制度文件只包括适用于甲、乙、丙、丁和戊的信息内容，则服务器只需获取目标信息树 M 对应的甲、乙、丙、丁和戊分别对应的关联信息树。

服务器根据制度描述信息生成信息节点，将拆分得到的多个制度子文件分别关联至信息节点。服务器将多个关联有不同制度子文件的信息节点分别添加至相同适用对象标识对应的关联信息树。例如，在上述举例中，将关联有制度子文件 A1 的信息节点添加至目标信息树 M 中适用对象标识甲对应的关联信息树 $M_{甲}$；将关联有制度子文件 A2 的信息节点添加至目标信息树 M 中适用对象标识乙对应的关联信息树 $M_{乙}$，如此类推。

当接收到第二终端发送的制度查询请求时，服务器获取适用对象标识对应的关联信息树。制度查询请求携带了适用对象标识和查询条件。服务器在关联信息树中查找满足查询条件的信息节点，获取与满足查询条件的信息节点关联的制度子文件，将制度子文件发送至第二终端。

本实施例中，在制度信息发布时，将记录来了适用于不同岗位的制度信息的制度文件拆分，将每个岗位需要执行或了解的制度条款挑选出来，满足不同岗位个性化需求，为不同岗位分别构建只包含相应岗位需求内容的关联信息树，且所有关联信息树的生成过程全自动进行，省时省力；后续用户只需基于适用于自己的关联信息树进行制度查询，也可以提高制度查询效率。

在其中一个实施例中，对制度文件进行拆分包括：计算制度文件的数据量，检测数据量是否超过阈值；当数据量超过阈值时，获取预设的目标数据量，根

据目标数据量确定制度文件的拆分位置；检测拆分位置是否位于相邻分隔符之间；当拆分位置位于一个分隔符处时，在拆分位置将制度文件拆分为多个中间文件；当拆分位置位于相邻分隔符之间时，在相邻分隔符中任意一个分隔符处将制度文件拆分为多个中间文件；按照预设的拆分规则，对多个中间文件进行拆分。

服务器计算制度文件的数据量，检测数据量是否超过阈值。该阈值可以是预先设定的，也可以是根据服务器的负载监测结果临时生成的。当数据量超过阈值时，服务器可以将制度文件预先拆分为多个数据量小的中间文件，再将中间文件分别拆分为多个制度子文件。具体的，服务器获取预设的目标数据量，根据目标数据量确定制度文件的拆分位置。目标数据量可以是预先设定的，也可以是根据对多个集群内其他服务器的负载监测结果临时生成的。例如，制度文件 A 的数据量为 720M，假设目标数据量为 80M，则将制度文件的第 80 M 大小的位置标记为第一个拆分位置，第 160 M 大小的位置标记为第二个拆分位置，以此类推。

服务器识别每个拆分位置是否位于相邻分隔符之间。当拆分位置位于一个分隔符所在的位置时，服务器在该拆分位置对制度文件进行拆分，得到该制度文件对应的多个中间文件。当拆分位置位于相邻分隔符之间时，服务器在相邻分隔符中任意一个分隔符处对相应制度文件进行拆分，即对该相邻分隔符中的前一个分隔符或后一个分隔符处进行拆分，得到制度文件对应的多个中间文件。服务器调用多线程按照上述方式将中间文件拆分为多个制度子文件，或者将中间文件发送至集群内其他服务器进行拆分，以提高文件拆分效率。将数据量较大的制度文件拆分为数据量较小的中间文件后传输至集群内其他服务器进行拆分，还可以提高数据传输效率，

本实施例中，对于数据量较大的制度文件进行两级拆分：其中，第一层级的拆分是根据数据量进行拆分，第二层级的拆分是根据预设的拆分维度进行拆分；将数据量较大的制度文件拆分为数据量较小的中间文件，可以并行将中间文件拆分为多个制度子文件，进而可以提高文件拆分效率。

应该理解的是，虽然图 2 和图 4 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它

的顺序执行。而且，图 2 和图 4 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 6 所示，提供了一种制度信息处理装置，包括：信息扩展模块 602、信息分类模块 604 和信息归档模块 606，其中：

信息扩展模块 602，用于监测终端发布的制度信息，对制度信息进行分词得到对应的原始词语集合；原始词语集合包括多个原始词语；对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；根据各个扩展词语集合形成制度信息对应的扩展制度信息集合。信息分类模块 604，用于将扩展制度信息集合输入预设的制度管理模型，得到制度信息对应的目标类别。信息归档模块 606，用于获取多个目标信息树分别对应的类别标注，筛选包含与目标类别对应类别标注的目标信息树，将制度信息添加至筛选得到的目标信息树。

在其中一个实施例中，制度信息包括制度描述信息；信息扩展模块 602 还用于检测制度描述信息是否包含类别信息；若包含，则根据类别信息将制度信息添加至相应的目标信息树；否则，对制度信息进行分词得到对应的原始词语集合。

在其中一个实施例中，该装置还包括模型训练模块 608，用于获取训练样本数据；训练样本数据包括多个样本制度信息及分别对应的类别标注；对各个样本制度信息进行分词和同义扩展处理，得到每个样本制度信息分别对应的扩展样本制度信息集合；根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，得到制度管理模型。

在其中一个实施例中，扩展样本制度信息集合包括多组扩展样本制度信息；模型训练模块 608 还用于获取特征项，计算特征项在一组扩展样本制度信息的词频权重；计算特征项在整个训练样本数据中的文档频率；根据词频权重和文档频率计算特征项对应的特征权重；根据特征权重选择特征项作为相应扩展样本制度信息的特征词；根据特征词提取各个扩展样本标准信息的特征。

在其中一个实施例中，制度信息包括制度描述信息和制度文件；信息归档

模块 606 还用于根据制度描述信息生成信息节点；检测筛选得到的目标信息树中是否已存在相同的信息节点；若不存在，则将信息节点添加至相应的目标信息树，将制度文件关联至信息节点。

在其中一个实施例中，制度信息包括制度描述信息及关联的制度文件；制度文件包括多个制度条款以及分别对应的适用对象标识；关联信息树具有对应的适用对象标识；信息归档模块 606 还用于对制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件；获取目标信息树对应的多个关联信息树；根据适用对象标识，将制度描述信息及制度子文件添加至相应的关联信息树。

在其中一个实施例中，信息归档模块606还用于计算制度文件的数据量，检测数据量是否超过阈值；当数据量超过阈值时，获取预设的目标数据量，根据目标数据量确定制度文件的拆分位置；检测拆分位置是否位于相邻分隔符之间；当拆分位置位于一个分隔符处时，在拆分位置将制度文件拆分为多个中间文件；当拆分位置位于相邻分隔符之间时，在相邻分隔符中任意一个分隔符处将制度文件拆分为多个中间文件；按照预设的拆分规则，对多个中间文件进行拆分。

关于制度信息处理装置的具体限定可以参见上文中对于制度信息处理方法的限定，在此不再赘述。上述制度信息处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在其中一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 7 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储制度信息。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时实现本申请任意一个实施例中提供的制度信息处理方法的步骤。

本领域技术人员可以理解，图 7 中示出的结构，仅仅是与本申请方案相关

的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一个或多个存储有计算机可读指令的非易失性存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现本申请任意一个实施例中提供的制度信息处理方法的步骤。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种制度信息处理方法，包括:

监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语;

对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合;

根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合;

将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别；及

获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

2、根据权利要求 1 所述的方法，其特征在于，所述制度信息包括制度描述信息；所述对所述制度信息进行分词得到对应的原始词语集合之前，所述方法还包括:

检测所述制度描述信息是否包含类别信息;

若包含，则根据所述类别信息将所述制度信息添加至相应的目标信息树;及

否则，对所述制度信息进行分词得到对应的原始词语集合。

3、根据权利要求 1 所述的方法，其特征在于，所述制度管理模型的生成，包括：获取训练样本数据；所述训练样本数据包括多个样本制度信息及分别对应的类别标注;

对各个所述样本制度信息进行分词和同义扩展处理，得到每个所述样本制度信息分别对应的扩展样本制度信息集合；及

根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，得到所述制度管理模型。

4、根据权利要求 3 所述的方法，其特征在于，所述扩展样本制度信息集合包括多组扩展样本制度信息；根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，包括:

获取特征项，计算所述特征项在一组所述扩展样本制度信息的词频权重;

计算所述特征项在整个训练样本数据中的文档频率;

根据所述词频权重和文档频率计算所述特征项对应的特征权重；

根据所述特征权重选择所述特征项作为相应扩展样本制度信息的特征词；及

根据所述特征词提取各个所述扩展样本标准信息的特征。

5、根据权利要求 1 所述的方法，其特征在于；所述制度信息包括制度描述信息和制度文件；所述将所述制度信息添加至筛选得到的目标信息树，包括：

根据所述制度描述信息生成信息节点；

检测筛选得到的目标信息树中是否已存在相同的信息节点；及

若不存在，则将所述信息节点添加至相应的目标信息树，将所述制度文件关联至所述信息节点。

6、根据权利要求 1 所述的方法，其特征在于，所述制度信息包括制度描述信息及关联的制度文件；所述制度文件包括多个制度条款以及分别对应的适用对象标识；所述关联信息树具有对应的适用对象标识；所述方法还包括：

对所述制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件；

获取所述目标信息树对应的多个关联信息树；及

根据所述适用对象标识，将所述制度描述信息及所述制度子文件添加至相应的关联信息树。

7、根据权利要求 6 所述的方法，其特征在于，所述对所述制度文件进行拆分，包括：

计算所述制度文件的数据量，检测所述数据量是否超过阈值；

当所述数据量超过阈值时，获取预设的目标数据量，根据所述目标数据量确定所述制度文件的拆分位置；

检测所述拆分位置是否位于相邻分隔符之间；

当所述拆分位置位于一个分隔符处时，在所述拆分位置将所述制度文件拆分为多个中间文件；

当所述拆分位置位于相邻分隔符之间时，在所述相邻分隔符中任意一个分隔符处将所述制度文件拆分为多个中间文件；及

按照预设的拆分规则，对多个所述中间文件进行拆分。

8、一种制度信息处理装置，包括:

信息扩展模块，用于监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语；对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合;

信息分类模块，用于将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别;

信息归档模块，用于获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

9、一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤:

监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语;

对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合;

根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合;

将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别；及

获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

10、根据权利要求 9 所述的计算机设备，其特征在于，所述制度信息包括制度描述信息；所述处理器执行所述计算机可读指令时还执行以下步骤:

检测所述制度描述信息是否包含类别信息;

若包含，则根据所述类别信息将所述制度信息添加至相应的目标信息树;及

否则，对所述制度信息进行分词得到对应的原始词语集合。

11、根据权利要求 9 所述的计算机设备，其特征在于，所述处理器执行

所述计算机可读指令时还执行以下步骤：获取训练样本数据；所述训练样本数据包括多个样本制度信息及分别对应的类别标注；

对各个所述样本制度信息进行分词和同义扩展处理，得到每个所述样本制度信息分别对应的扩展样本制度信息集合；及

根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，得到所述制度管理模型。

12、根据权利要求 11 所述的计算机设备，其特征在于，所述扩展样本制度信息集合包括多组扩展样本制度信息；所述处理器执行所述计算机可读指令时还执行以下步骤：

获取特征项，计算所述特征项在一组所述扩展样本制度信息的词频权重；

计算所述特征项在整个训练样本数据中的文档频率；

根据所述词频权重和文档频率计算所述特征项对应的特征权重；

根据所述特征权重选择所述特征项作为相应扩展样本制度信息的特征词；及

根据所述特征词提取各个所述扩展样本标准信息的特征。

13、根据权利要求 9 所述的计算机设备，其特征在于，所述制度信息包括制度描述信息和制度文件；所述处理器执行所述计算机可读指令时还执行以下步骤：

根据所述制度描述信息生成信息节点；

检测筛选得到的目标信息树中是否已存在相同的信息节点；及

若不存在，则将所述信息节点添加至相应的目标信息树，将所述制度文件关联至所述信息节点。

14、根据权利要求 9 所述的计算机设备，其特征在于，所述制度信息包括制度描述信息及关联的制度文件；所述制度文件包括多个制度条款以及分别对应的适用对象标识；所述关联信息树具有对应的适用对象标识；所述处理器执行所述计算机可读指令时还执行以下步骤：

对所述制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件；

获取所述目标信息树对应的多个关联信息树；及

根据所述适用对象标识，将所述制度描述信息及所述制度子文件添加至相应的关联信息树。

15、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

监测终端发布的制度信息，对所述制度信息进行分词得到对应的原始词语集合；所述原始词语集合包括多个原始词语；

对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合；

根据各个扩展词语集合形成所述制度信息对应的扩展制度信息集合；

将所述扩展制度信息集合输入预设的制度管理模型，得到所述制度信息对应的目标类别；及

获取多个目标信息树分别对应的类别标注，筛选包含与所述目标类别对应类别标注的目标信息树，将所述制度信息添加至筛选得到的目标信息树。

16、根据权利要求 15 所述的存储介质，其特征在于，所述制度信息包括制度描述信息；所述计算机可读指令被所述处理器执行时还执行以下步骤：

检测所述制度描述信息是否包含类别信息；

若包含，则根据所述类别信息将所述制度信息添加至相应的目标信息树；及

否则，对所述制度信息进行分词得到对应的原始词语集合。

17、根据权利要求 15 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：获取训练样本数据；所述训练样本数据包括多个样本制度信息及分别对应的类别标注；

对各个所述样本制度信息进行分词和同义扩展处理，得到每个所述样本制度信息分别对应的扩展样本制度信息集合；及

根据各个扩展样本制度信息集合和对应的类别标注，通过支持向量机算法对初始的制度管理模型进行训练，得到所述制度管理模型。

18、根据权利要求 17 所述的存储介质，其特征在于，所述扩展样本制度信息集合包括多组扩展样本制度信息；所述计算机可读指令被所述处理器执行时还执行以下步骤：

获取特征项，计算所述特征项在一组所述扩展样本制度信息的词频权重；

计算所述特征项在整个训练样本数据中的文档频率；

根据所述词频权重和文档频率计算所述特征项对应的特征权重；

根据所述特征权重选择所述特征项作为相应扩展样本制度信息的特征词；及

根据所述特征词提取各个所述扩展样本标准信息的特征。

19、根据权利要求 15 所述的存储介质，其特征在于，所述制度信息包括制度描述信息和制度文件；所述计算机可读指令被所述处理器执行时还执行以下步骤：

根据所述制度描述信息生成信息节点；

检测筛选得到的目标信息树中是否已存在相同的信息节点；及

若不存在，则将所述信息节点添加至相应的目标信息树，将所述制度文件关联至所述信息节点。

20、根据权利要求 15 所述的存储介质，其特征在于，所述制度信息包括制度描述信息及关联的制度文件；所述制度文件包括多个制度条款以及分别对应的适用对象标识；所述关联信息树具有对应的适用对象标识；所述计算机可读指令被所述处理器执行时还执行以下步骤：

对所述制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件；

获取所述目标信息树对应的多个关联信息树；及

根据所述适用对象标识，将所述制度描述信息及所述制度子文件添加至相应的关联信息树。

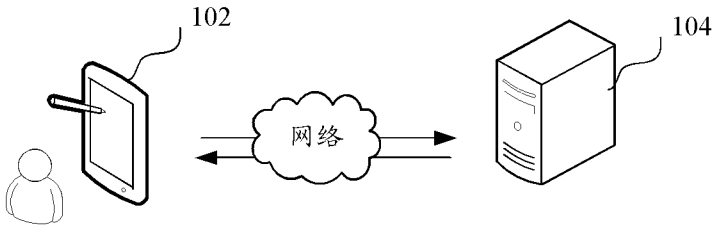

102
104
网络

图 1

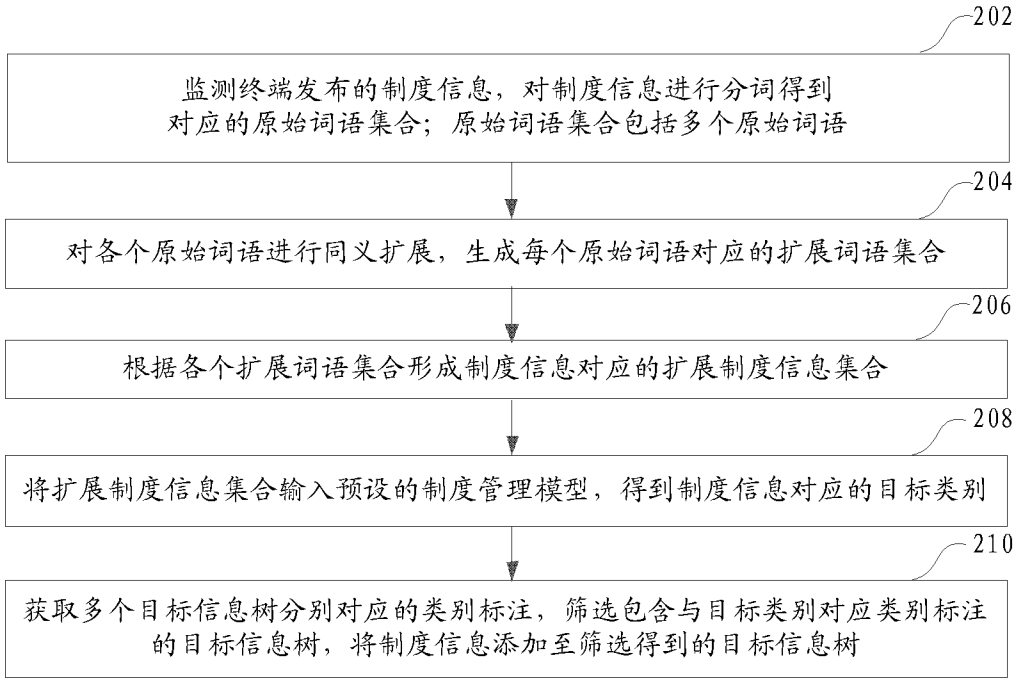

202
监测终端发布的制度信息，对制度信息进行分词得到对应的原始词语集合；原始词语集合包括多个原始词语
204
对各个原始词语进行同义扩展，生成每个原始词语对应的扩展词语集合
206
根据各个扩展词语集合形成制度信息对应的扩展制度信息集合
208
将扩展制度信息集合输入预设的制度管理模型，得到制度信息对应的目标类别
210
获取多个目标信息树分别对应的类别标注，筛选包含与目标类别对应类别标注的目标信息树，将制度信息添加至筛选得到的目标信息树

图 2

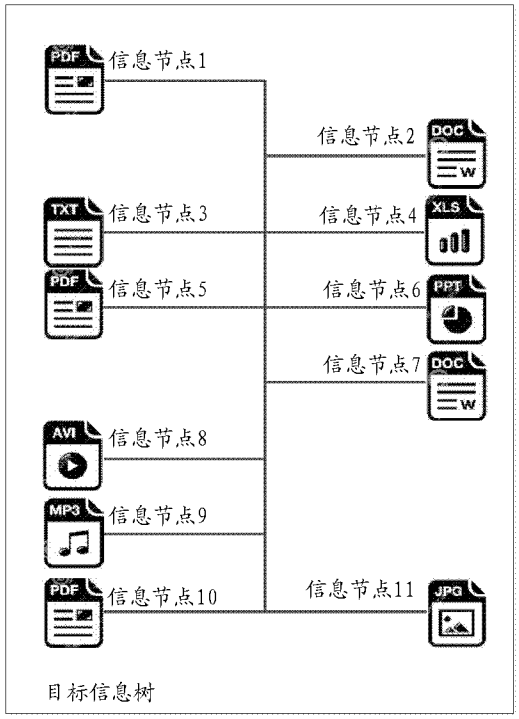

信息节点1
信息节点2
信息节点3
信息节点4
信息节点5
信息节点6
信息节点7
信息节点8
信息节点9
信息节点10
信息节点11
目标信息树

图 3

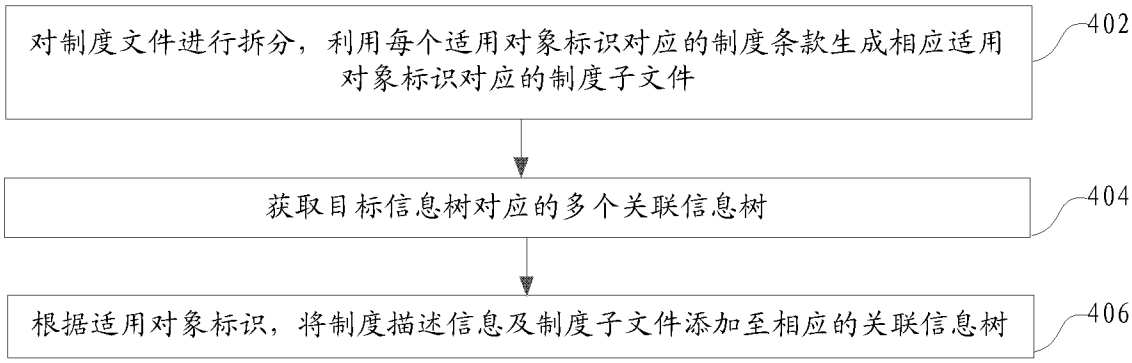

对制度文件进行拆分，利用每个适用对象标识对应的制度条款生成相应适用对象标识对应的制度子文件
402
获取目标信息树对应的多个关联信息树
404
根据适用对象标识，将制度描述信息及制度子文件添加至相应的关联信息树
406

图 4

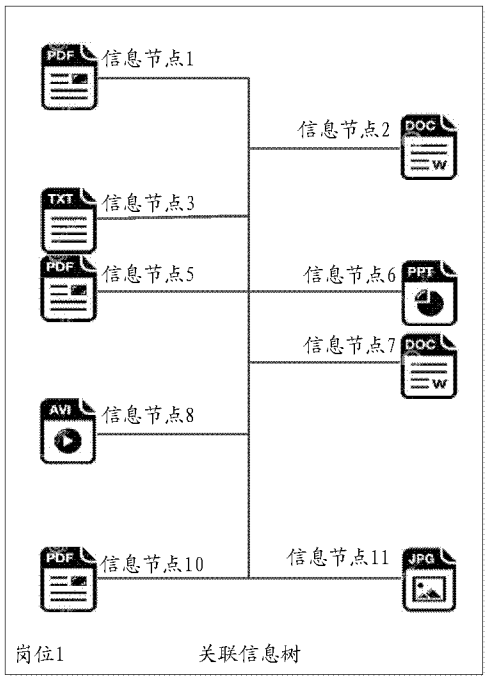

信息节点1
信息节点2
信息节点3
信息节点5
信息节点6
信息节点7
信息节点8
信息节点10
信息节点11
岗位1
关联信息树

图 5

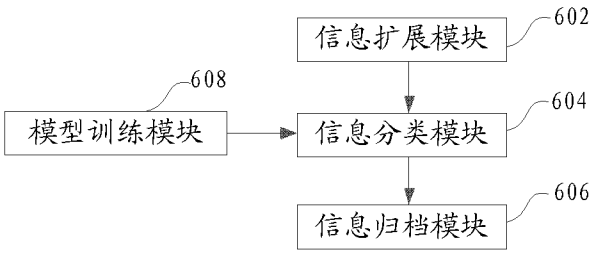

602
信息扩展模块
608
模型训练模块
604
信息分类模块
606
信息归档模块

图 6

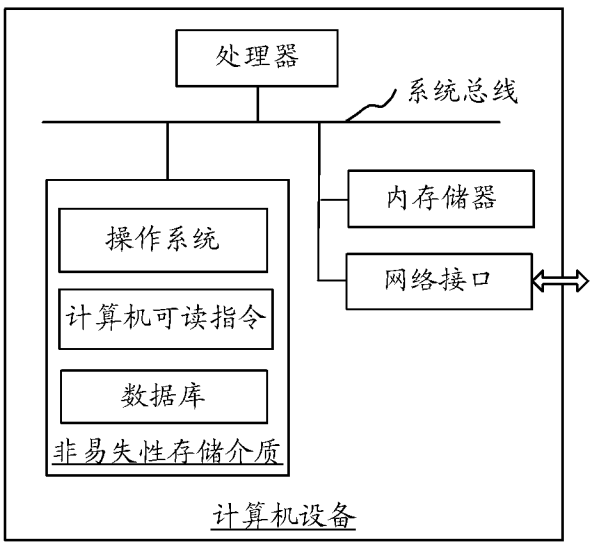

处理器
系统总线
内存储器
操作系统
网络接口
计算机可读指令
数据库
非易失性存储介质
计算机设备

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/095549

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27(2006.01)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)

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)

CNABS; DWPI; SIPOABS; CNKI: 信息, 数据, 分类, 类别, 词, 扩展, 模型, 树, 训练, 学习, 权重, 频率, 拆分; information, data, classif+, sort, kind, word, vocable, extend+, model, tree, train+, learn+, weight, frequency, split+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107436875 A (HUAWEI TECHNOLOGIES CO., LTD.) 05 December 2017 (2017-12-05) claims 1-17, and description, paragraphs [0007]-[0048] and [0190]	1, 2, 5-10, 13-16, 19, 20
Y	CN 107436875 A (HUAWEI TECHNOLOGIES CO., LTD.) 05 December 2017 (2017-12-05) claims 1-17, and description, paragraphs [0007]-[0048] and [0190]	3, 4, 11, 12, 17, 18
Y	CN 104978328 A (BEIJING QIHOO TECHNOLOGY CO., LTD. ET AL.) 14 October 2015 (2015-10-14) claims 1-10, and description, paragraphs [0030]-[0059]	3, 4, 11, 12, 17, 18
A	CN 107862046 A (NINGBO AISINO SPACEFLIGHT INFORMATION CO., LTD.) 30 March 2018 (2018-03-30) 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	Date of mailing of the international search report
17 December 2018	**15 January 2019**

Name and mailing address of the ISA/CN	Authorized officer
National Intellectual Property Administration, PRC (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/095549

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 107436875 A	05 December 2017	WO 2017202125 A1	30 November 2017
CN 104978328 A	14 October 2015	None	
CN 107862046 A	30 March 2018	None	

国际检索报告

国际申请号
PCT/CN2018/095549

A. 主题的分类

G06F 17/27(2006.01)i

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

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））

CNABS;DWPI;SIPOABS;CNKI:信息，数据，分类，类别，词，扩展，模型，树，训练，学习，权重，频率，拆分;information,data,classif+,sort,kind,word,vocable,extend+,model,tree,train+,learn+,weight,frequency,split+

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
X	CN 107436875 A (华为技术有限公司) 2017年 12月 5日 (2017-12-05) 权利要求1-17、说明书第【0007】-【0048】，【0190】段	1-2，5-10，13-16，19-20
Y	CN 107436875 A (华为技术有限公司) 2017年 12月 5日 (2017-12-05) 权利要求1-17、说明书第【0007】-【0048】，【0190】段	3-4，11-12，17-18
Y	CN 104978328 A (北京奇虎科技有限公司 等) 2015年 10月 14日 (2015-10-14) 权利要求1-10、说明书第【0030】-【0059】段	3-4，11-12，17-18
A	CN 107862046 A (宁波爱信诺航天信息有限公司) 2018年 3月 30日 (2018-03-30) 全文	1-20

☐ 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，但为了理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	国际检索报告邮寄日期
2018年 12月 17日	2019年 1月 15日

ISA/CN的名称和邮寄地址	受权官员
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	崔哲
传真号 (86-10)62019451	电话号码 62411849

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/095549

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 107436875 A	2017年 12月 5日	WO 2017202125 A1	2017年 11月 30日
CN 104978328 A	2015年 10月 14日	无	
CN 107862046 A	2018年 3月 30日	无	