

(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
2017 年 8 月 3 日 (03.08.2017)



(10) 国际公布号
WO 2017/128997 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/071409
- (22) 国际申请日: 2017 年 1 月 17 日 (17.01.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610055298.5 2016 年 1 月 27 日 (27.01.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 胡于响 (HU, Yuxiang) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, 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)。

(54) Title: SERVICE PROCESSING METHOD, AND DATA PROCESSING METHOD AND DEVICE

(54) 发明名称: 业务处理方法、数据处理方法及装置

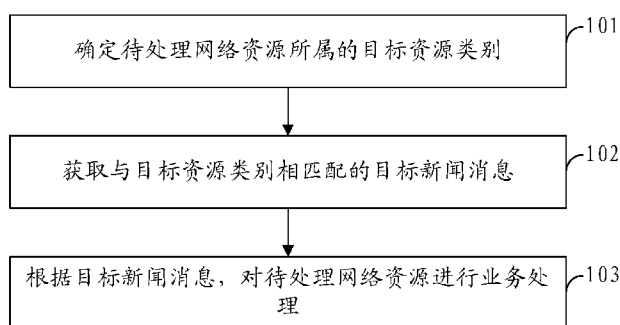


图 1

- 101 Determining a target resource category to which a network resource to be processed belongs
- 102 Acquiring a target news message matching the target resource category
- 103 According to the target news message, performing service processing on the network resource to be processed

(57) Abstract: Provided are a service processing method, and a data processing method and device. The service processing method comprises: determining a target resource category to which a network resource to be processed belongs; acquiring a target news message matching the target resource category; and according to the target news message, performing service processing on the network resource to be processed. Provided is a new service processing method. The present application can improve the quality of service processing, and enrich service processing modes.

(57) 摘要: 本申请提供一种业务处理方法、数据处理方法及装置。业务处理方法包括: 确定待处理网络资源所属的目标资源类别; 获取与目标资源类别相匹配的目标新闻消息; 根据所述目标新闻消息, 对待处理网络资源进行业务处理。本申请提供一种新的业务处理方法, 可以提高业务处理质量, 丰富业务处理方式。

WO 2017/128997 A1

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

业务处理方法、数据处理方法及装置

【技术领域】

本申请涉及互联网技术领域，尤其涉及一种业务处理方法、数据处理方法及装置。

【背景技术】

随着互联网技术的发展，网络资源越来越多。依赖于网络资源的业务处理也越来越多，例如与网络资源相关的信息推送、网络资源的上传/下载、网络资源的获取、以及网络资源管理等。

在现有业务处理过程中，主要依赖网络资源自身的属性信息。在某些情况下，业务处理过程可能会收到外界信息的影响，例如，在电子商务领域，一些商品的销售量往往会受热点新闻和资讯的影响。因此，现有业务处理方式的比较单一，处理效果不佳，因此需要一种新的业务处理方法。

【发明内容】

本申请的多个方面提供一种业务处理方法、数据处理方法及装置，用以提供一种新的业务处理方法，提高业务处理质量，丰富业务处理方式。

本申请的一方面，提供一种业务处理方法，包括：

确定待处理网络资源所属的目标资源类别；

获取与所述目标资源类别相匹配的目标新闻消息；

根据所述目标新闻消息，对所述待处理网络资源进行业务处理。

本申请的另一方面，提供一种数据处理方法，包括：

按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算所述新闻消息与资源类别库中各资源类别之间的相似度；

确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别；

建立所述新闻消息和所述确定的资源类别之间的匹配关系。

本申请的又一方面，提供一种业务处理方法，包括：

按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

确定与所述新闻消息相匹配的目标资源类别；

对所述目标资源类别下的网络资源进行业务处理。

本申请的又一方面，提供一种业务处理装置，包括：

第一确定模块，用于确定待处理网络资源所属的目标资源类别；

获取模块，用于获取与所述目标资源类别相匹配的目标新闻消息；

业务模块，用于根据所述目标新闻消息，对所述待处理网络资源进行业务处理。

本申请的又一方面，提供一种数据处理装置，包括：

抓取模块，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算模块，用于计算所述新闻消息与资源类别库中各资源类别之间的相似度；

确定模块，用于确定与所述新闻消息之间的相似度满足预设第一相似条件的资源类别；

建立模块，用于建立所述新闻消息和所述确定的资源类别之间的匹配关系。

本申请的又一方面，提供一种业务处理装置，包括：

抓取模块，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

确定模块，用于确定与所述新闻消息相匹配的目标资源类别；

业务模块，用于对所述目标资源类别下的网络资源进行业务处理。

在本申请中，确定待处理网络资源所属的目标资源类别，获取与该目标资源类别相匹配的目标新闻消息，根据目标新闻消息，对待处理资源网络进行业务处理，或者抓取新闻消息，确定与新闻消息相匹配的目标资源类别，根据目标资源类别进行业务处理，提供一种基于新闻消息与资源类别之间的匹配关系的业务处理方法，充分发挥新闻消息对业务处理过程的影响，提高

业务处理精度，同时丰富业务处理方式。

【附图说明】

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，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 为本申请一实施例提供的业务处理方法的流程示意图；

图 2a 为本申请另一实施例提供的数据处理方法的流程示意图；

图 2b 和图 2c 为本申请又一实施例提供的用于执行图 2a 所示方法的系统架构示意图；

图 2d 为本申请又一实施例列举的新闻消息与资源类目之间匹配关系的示意图；

图 3 为本申请又一实施例提供的业务处理方法的流程示意图；

图 4 为本申请又一实施例提供的业务处理装置的结构示意图；

图 5 为本申请又一实施例提供的业务处理装置的结构示意图；

图 6 为本申请又一实施例提供的数据处理装置的结构示意图；

图 7 为本申请又一实施例提供的数据处理装置的结构示意图；

图 8 为本申请又一实施例提供的业务处理装置的结构示意图。

【具体实施方式】

为使本申请实施例的目的、技术方案和优点更加清楚，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做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。

图 1 为本申请一实施例提供的业务处理方法的流程示意图。如图 1 所示，

- 101、确定待处理网络资源所属的目标资源类别。
- 102、获取与目标资源类别相匹配的目标新闻消息。
- 103、根据目标新闻消息，对待处理网络资源进行业务处理。

本实施例提供一种业务处理方法，可由业务处理装置来执行，用以实现新的业务处理流程，提高业务处理质量，丰富业务处理方式。

经过本申请发明人的分析研究，发现新闻消息与网络资源、以及依赖于网络资源的业务有着紧密的联系。本申请发明人最为直接的发现为：在电子商务领域，一些商品的销售量往往会受热点新闻和资讯的影响。举例说明，近期天津爆炸案相关的热点新闻，引发人们去注意消防安全及环境污染，从而促使灭火器、口罩、消毒水等这样的商品销售量上升；有关“柴静发表‘穹顶之下’演说”的热点新闻，引发人们去注意身边的空气质量，从而促使防雾霾口罩等商品销售量上升；“成都女司机被打”的热点新闻中提到了行车记录仪，引起了大家对行车记录仪的重视，行车记录仪相关产品也迎来销售高峰。

基于上述考虑，本申请发明人提供一种新的业务处理方法，其主要原理是：基于资源类别与新闻消息之间的匹配关系进行业务处理。在本申请提供的业务处理方法中，涉及新闻消息、业务处理、以及网络资源。为便于描述，将本申请业务处理方法涉及的网络资源称为待处理网络资源，将网络资源所属的资源类别称为目标资源类别，将与目标资源类别相匹配的新闻消息称为目标新闻消息。

首先说明，本申请实施例不限制新闻消息的内容，例如可以包括新闻事件、热点话题、人物动态、产品资讯等中的至少一种；另外，也不限制新闻消息的实现格式，例如可以包括文本、图片、视频等中的至少一种。

另外，本申请实施例中的资源类别是指网络资源所属的类别。本申请实施例不限定网络资源的类型。在不同应用场景中，网络资源会有所不同，网络资源所属的类别也会有所不同。举例说明：

在电子商务领域，网络资源可以是卖家提供的各种商品、服务等，相应的，资源类别可以是网络资源所属的类目，例如女装、男装、鞋子、生活、学习、运动、户外、母婴等。值得说明的是，本申请实施例并不限制类目等级，也就是说，在本申请实施例中，资源类别可以包括各种等级的类目。

基于上述介绍，本实施例的业务处理方法具体包括：

确定待处理网络资源所属的资源类别作为目标资源类别；然后，获取与目标资源类别相匹配的新闻消息作为目标新闻消息，之后，根据该目标新闻消息，对待处理网络资源进行业务处理。

值得说明的是，根据应用场景的不同，根据目标新闻消息，对待处理网络资源进行业务处理的具体流程也会有所不同。下面以电子商务领域中的部分业务场景为例进行举例说明，对本领域技术人员来说，在下述举例说明的基础上，可以实现其他应用场景中根据目标新闻消息对待处理网络资源进行业务处理的流程。

在电子商务领域中，电商平台向用户（这里的用户可以是B类用户，也可以是C类用户）进行商品推荐也是一种较为常见的业务场景。由于热门新闻和资讯会影响商品价格和热度，所以可以采用本实施例提供的方法可以向用户推荐一些与热门新闻和资讯有关的商品。这里的B类用户是指B类贸易场景中的用户，这类用户购买商品不是用于自己消费，而是用于再次交易，例如售卖或者加工生产。这里的C类用户是指C类贸易场景中的用户，这类用户是一般的消费者，其购买商品主要是用于自身消费。具体的，确定待推荐商品所属的类目作为目标类目，获取与目标类目相匹配的新闻消息作为目标新闻消息，根据该目标新闻消息，对待推荐商品进行推荐处理。

其中，上述根据该目标新闻消息，对待推荐商品进行推荐处理包括但不限于以下处理：

根据该目标新闻消息，确定该待推荐商品是否具有推荐价值，也就是确定是否向用户推荐该待推荐商品；例如，防雾霾口罩所属类目与有关“柴静发表‘穹顶之下’演说”这一热点新闻相匹配，根据有关“柴静发表‘穹顶之下’演说”这一热点新闻，可以确定防雾霾口罩具有推荐价值，因此可以向用户推荐防雾霾口罩；

进一步，若确定向用户推荐该待推荐商品，还可以确定待推荐商品的品牌（即推荐哪些品牌的商品）、产地（推荐哪些产地的商品）、价格区间（推荐位于哪些价格区间的商品）、卖家信息（推荐哪些卖家的商品）、图片以及推荐所使用的文字信息等中的至少一项。

另外，在电子商务领域中，电商平台向用户（这里是指卖家）提供采购

决策也是一种较为主要的业务场景。由于热门新闻和资讯会影响商品价格和热度，所以可以采用本实施例提供的方法可以向卖家提供更加精确的采购决策。具体的，确定各种商品所属的类目作为目标类目，获取与目标类目相匹配的新闻消息作为目标新闻消息，根据该目标新闻消息，对各种商品生成用户的采购策略。

其中，上述根据该目标新闻消息，对各种商品生成用户的采购策略包括但不限于以下处理：

对于每种商品，根据目标新闻消息，确定该商品对用户来说是否具有采购价值，也就是确定用户是否需要采购该商品；例如，防雾霾口罩所属类目与有关“柴静发表‘穹顶之下’演说”这一热点新闻相匹配，根据有关“柴静发表‘穹顶之下’演说”这一热点新闻，可以确定防雾霾口罩近期销量将大幅上升，确定防雾霾口罩具有采购价值，因此可以确定采购防雾霾口罩；

进一步，若确定用户需要采购该商品，还可以确定采购数量、采购价格、采购周期、采购商家等中的至少一项。

由上述可见，本实施例首先确定待处理网络资源所属的目标资源类别，获取与该目标资源类别相匹配的目标新闻消息，根据目标新闻消息，对待处理资源网络进行业务处理，提供一种基于新闻消息与资源类别之间的匹配关系的业务处理方法，充分发挥新闻消息对业务处理过程的影响，提高业务处理精度，同时丰富业务处理方式。

在一可选实施方式中，可以预先建立资源类别与新闻消息之间的匹配关系。基于此，上述步骤 102，即获取与目标资源类别相匹配的目标新闻消息具体为：根据目标资源类别，查询预先建立的资源类别与新闻消息之间的匹配关系，以获取与目标资源类别相匹配的新闻消息作为目标新闻消息。

其中，图 2a 为本申请另一实施例提供的数据处理方法的流程示意图。该数据处理方法用于预先建立资源类别与新闻消息之间的匹配关系。例如，在上述实施方式中，可以采用图 2a 所示方法预先建立资源类别与新闻消息之间的匹配关系，然后根据目标资源类别，查询资源类别与新闻消息之间的匹配关系，获得与该目标资源类别相匹配的目标新闻消息。值得说明的是，采用图 2a 所述方法建立的资源类别与新闻消息之间的匹配关系可以应用于各种需要该匹配关系的应用场景中，并不仅仅适用于上述实施方式。如图 2a 所示，

该方法包括：

- 201、按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。
- 202、计算上述新闻消息与资源类别库中各资源类别之间的相似度。
- 203、确定与新闻消息之间的相似度满足预设第一相似度条件的资源类别。
- 204、建立新闻消息和所确定的资源类别之间的匹配关系。

图 2a 所示方法流程可以采用但不限于图 2b 和 2c 系统架构来执行。具体的，图 2b 所示的抓取引擎可以按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；抓取引擎抓取到的新闻消息可以存储在图 2b 所示的数据存储系统，所述数据存储系统可采用 mysql 关系型数据库实现，但不限于此。图 2c 所示信息提取平台进行信息提取，并根据提取出的信息完成新闻消息和所确定的资源类别之间的匹配关系的建立。

其中，步骤 201 中的抓取周期可以根据应用场景适应性设置，例如可以是一天、一周、三天、五天等。另外，考虑到网络平台上新闻消息的数量较多，并且这些新闻消息的价值会随着时间的增长而递减，所以在本实施例中，预先设定要求，具体抓取满足预设要求的新闻消息，这样可以降低新闻消息的数量，提高处理效率。所述预设要求可是热度大于指定热度阈值（这样可以获取热点新闻消息），或者是出现时间晚于指定时间（这样可以获取近期出现的新闻消息）。

例如，抓取引擎可以利用爬虫，将各大新闻网站（例如新浪网，其网址为 www.sina.com.cn，搜狐网，其网址为 www.sohu.com 等）上的热点新闻抓取下来。所谓热点新闻也就是热度比较高的新闻消息，例如具体可能是各大新闻网站上比较靠前的新闻消息，如头条消息等。优选地，这里的爬虫可采用 Jsoup 定向抓取技术，但不限于此。

对于抓取到的新闻消息，计算该新闻消息与资源类别库中各资源类别之间的相似度，进而确定与该新闻消息之间的相似度满足预设第一相似度条件的资源类别作为与该新闻消息相匹配的资源类别，然后建立该新闻消息与所确定的资源类别之间的匹配关系。

值得说明的是，在每个抓取周期内一般会抓取到多个新闻消息，对每个新闻消息均采用上述方法进行处理。另外，随着抓取周期的递增，可以建立大量新闻消息与资源类别之间的匹配关系。

可选的，可以将上述新闻消息与资源类别之间的匹配关系存储到数据存储系统中，但并不限于数据库。

进一步，上述步骤 202 的一种实现方式包括：

根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词；

对各资源类别分别进行分词处理，以获得各资源类别的关键词；

根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度。

进一步，抓取引擎可以细分为引擎管理模块、新闻抓取模块、评论抓取模块和数据接口模块。其中，引擎管理模块负责管理网络上的 URL（记为 URL 管理）以及管理需要抓取的 URL（简称为抓取点管理）。

新闻消息用于描述一个事实，新闻消息的正文、标题能够表达出新闻消息的主要含义。具体的，新闻抓取模块可以抓取新闻消息，并通过数据接口模块将抓取到的新闻消息存储到数据存储系统中的新闻信息表中。而新闻消息的评论信息则能体现网络用户（可简称为网友）的关注点。例如，在《成都女司机被打事件谁之过？》这个新闻消息中，全文没有提到行车记录仪，单从新闻消息本身是无法挖掘到行车记录仪这个信息，但是在下面的网友评论中，却有许多人都提到了行车记录仪的重要性。具体的，评论抓取模块可以抓取新闻消息的评论信息，并通过数据接口模块将抓取到评论信息存储到数据存储系统中的新闻评论表中。

基于上述，信息提取平台具体可以从数据存储系统中获取新闻消息的正文、标题和评论信息中的至少一类信息；然后，对新闻消息的正文、标题和评论信息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；将正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得新闻消息的关键词。

进一步，图 2c 所示信息提取平台可以细分为主题词提取模块、标题分词模块以及合并和去重模块。

可选的，对新闻消息的正文来说，由于其信息量较多，故对其进行关键词提取处理的方式可以为：可由主题词提取模块对其进行主题词提取处理；对新闻消息的标题来说，由于其相对简单，故可由标题分词模块对其进行关

关键词提取处理的方式可以为：对其进行分词处理；对新闻消息的评论信息来说，由于其信息量较多，故对其进行关键词提取处理的方式可以为：可由主题词提取模块对其进行主题词提取处理。

这里之所以要进行去重去处理，是考虑到爬虫从各大新闻网站上抓取热点新闻可能存在重复。例如天津爆炸案，一时间是各大新闻网站的头条，所以很可能会从不同新闻网站上抓取到相同的新闻消息，所以可以出现重复或相似度极高的关键词，因此需要将重复或相似度极高（例如大于一定阈值）的关键词合并为一个。

可选的，这里的去重处理具体可以采用聚类算法实现。具体的，对正文关键词、标题关键词和评论关键词中的至少一种进行聚类处理，将聚为一类的关键词用其中一个关键词代替。例如，可以采用向量空间模型描述这些关键词，采用凝聚层次聚类算法对这些关键词进行聚类，从而将相似关键词归到同一类中。例如，对于天津爆炸案相关的新闻消息，可以提取出来的关键词包括“爆炸”、“火灾”、“消防员”、“滨海新区”、“危化物”、“环境污染”、“死伤”等，其中“火灾”、“爆炸”和“消防员”这几个关键词均与“灭火器”相关子类别的相似度极高，因此需要将这几个关键词聚为一类，然后从中选一个关键词代表该聚类中的所有关键词。

优选的，可以同时采用新闻消息的正文、标题和评论信息，获得新闻消息的关键词。则一种同时采用新闻消息的正文、标题和评论信息，获得新闻消息的关键词的具体实施方式包括：提取处理，过滤处理，以及合并和去重处理。

提取处理是指：对新闻消息的正文和评论信息分别进行主题词提取处理，获得正文关键词和评论关键词，对新闻消息的标题进行分词处理，以获得标题关键词。

可选的，为了便于对这三类信息分别进行处理，可以用两张表对这三类信息进行存储，例如，将新闻消息的正文和标题存储在图 2b 所示新闻信息表中，将新闻消息的评论信息存储在图 2b 所示新闻评论表中，以便于分别进行处理。

在对新闻消息的评论信息进行主题词提取处理的过程中，可以采用 TF-IDF 模型，提取出网友的关注点作为评论关键词。例如，在成都女司机被

打新闻消息中，网友的评论中大量出现行车记录仪，通过 TF-IDF 模型可以很快挖掘出行车记录仪这个评论关键词。

过滤处理是指：对正文关键词、标题关键词和评论关键词分别进行过滤，以去除其中的停用词、人名、地名、时间等词。例如对《工程院院士：“阅兵蓝”后京津冀污染物显著上升》这篇新闻的标题进行分词处理得到的标题关键词包括“工程院院士”、“阅兵”、“蓝”、“后”、“京津冀”、“污染物”、“显著”和“上升”等词，这里的“后”属于停用词，将其去除。

合并和去重处理是指：将过滤后的正文关键词、标题关键词和评论关键词进行合并和去重处理，以获得新闻消息的关键词。

在获得新闻消息的关键词之后，可以根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度。一种根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度的实施方式包括：

获取新闻消息的关键词的词向量和各资源类别的关键词的词向量；

根据新闻消息的关键词的词向量与各资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度。

例如，在实际应用中，可以采用 Word2Vec 模型，计算新闻消息与各资源类别之间的相似度。Word2Vec 模型需要使用语料库，在本实施例中，该语料库可由大量与网络资源相关的新闻消息、网络资源提供商提供的网络资源及其详情、新闻消息的评论信息、资源类别信息等组成。

可选的，考虑到新闻消息的关键词可能有多个，每个资源类目的关键词也可能有多个，则将新闻消息的关键词的个数记为 n ，将资源类目的关键词的个数记为 m ，其中， n 和 m 是大于 1 的自然数。基于此，一种根据新闻消息的关键词的词向量与各资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度的实施方式包括：

对于 n 个关键词中的每个关键词，分别计算该关键词的词向量与 m 个关键词中的每个关键词的词向量之间的相似度，获得 $n*m$ 个相似度；

对 $n*m$ 个相似度求平均，将 $n*m$ 个相似度的平均值作为 n 个关键词所对应的新闻消息和 m 个关键词所对应的资源类目之间的相似度。

按照上述方式，可以计算出一新闻消息与各资源类目之间的相似度，进而从中选择满足预设第一相似度条件的资源类目，作为与该新闻消息相匹配的资源类别。以电子商务领域为例，一种新闻消息与资源类目之间的匹配关系如图 2d 所示。在图 2d 中，左侧为“柴静发表‘穹顶之下’演说”这一热点新闻，右侧是与该新闻消息相匹配的防雾霾口罩所属类目。

在另一可选实施方式中，上述步骤 102，即获取与目标资源类别相匹配的目标新闻消息具体为：计算新闻语料库中各新闻消息与目标资源类别之间的相似度；获取与目标资源类别之间的相似度满足预设第二相似度条件的新闻消息作为目标新闻消息。

其中，上述计算新闻语料库中各新闻消息与所述目标资源类别之间的相似度的一种实施方式包括：

对目标资源类别进行分词处理，以获得目标资源类别的关键词；

对于每个新闻消息，根据该新闻消息的正文、标题和评论信息中的至少一类信息，获取该新闻消息的关键词，根据新闻消息的关键词与目标资源类别的关键词，计算新闻消息与目标资源类别之间的相似度。

值得说明的是，该实施方式与上述步骤 202 的实施方式相类似，其各步骤的具体实现可参见上述步骤 202 的具体实施方式中的相应描述，在此不再赘述。

相应的，上述根据新闻消息的关键词与目标资源类别的关键词，计算新闻消息与目标资源类别之间的相似度包括：

获取新闻消息的关键词的词向量和目标资源类别的关键词的词向量；

根据新闻消息的关键词的词向量与目标资源类别的关键词的词向量，计算新闻消息和目标资源类别之间的相似度。

例如，在实际应用中，可以采用 Word2Vec 模型，计算新闻消息与目标资源类别之间的相似度。

可选的，考虑到新闻消息的关键词可能有多个，目标资源类目的关键词也可能有多个，则将新闻消息的关键词的个数记为 1，将目标资源类目的关键词的个数记为 k，其中，1 和 k 是大于 1 的自然数。基于此，一种根据新闻

消息的关键词的词向量与目标资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度的实施方式包括：

对于1个关键词中的每个关键词，分别计算该关键词的词向量与k个关键词中的每个关键词的词向量之间的相似度，获得 $1*k$ 个相似度；

对 $1*k$ 个相似度求平均，将 $1*k$ 个相似度的平均值作为1个关键词所对应的新闻消息和k个关键词所对应的目标资源类目之间的相似度。

按照上述方式，可以计算出新闻语料库中各新闻消息与目标资源类目之间的相似度，进而从中选择满足预设第二相似度条件的新闻消息，作为与目标资源类目匹配的新闻消息。

进一步可选的，在获得新闻消息与资源类别之间的匹配关系之后，可以将该匹配关系存储到图2c所示的数据存储系统中，具体可以存储到图2c所示数据存储系统中的匹配结果列表中。

以上方法实施例是从网络资源的角度出发，找到相匹配的新闻消息，进而基于相匹配的新闻消息对网络资源进行业务处理。以下方法实施例将从新闻消息的角度出发，找到与新闻消息相匹配的资源类别，然后对与所述新闻消息相匹配的资源类别下的网络资源进行业务处理。

图3为本申请又一实施例提供的业务处理方法的流程示意图。如图3所示，该方法包括：

301、按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。

302、确定与新闻消息相匹配的目标资源类别。

303、对目标资源类别下的网络资源进行业务处理。

本实施例提供一种业务处理方法，可由业务处理装置来执行，用以实现新的业务处理流程，提高业务处理质量，丰富业务处理方式。

经过本申请发明人的分析研究，发现新闻消息与网络资源、以及依赖于网络资源的业务有着紧密的联系。本申请发明人最为直接的发现为：在电子商务领域，一些商品的销售量往往会受热点新闻和资讯的影响。举例说明，近期天津爆炸案相关的热点新闻，引发人们去注意消防安全及环境污染，从而促使灭火器、口罩、消毒水等这样的商品销售量上升；有关“柴静发表‘穹顶之下’演说”的热点新闻，引发人们去注意身边的空气质量，从而促使防雾

霾口罩等商品销售量上升；“成都女司机被打”的热点新闻中提到了行车记录仪，引起了大家对行车记录仪的重视，行车记录仪相关产品也迎来销售高峰。

基于上述考虑，本申请发明人提供一种新的业务处理方法，其主要原理是：基于资源类别与新闻消息之间的匹配关系进行业务处理。为便于描述，将本申请业务处理方法中与新闻消息相匹配的资源类别称为目标资源类别。

首先说明，本申请实施例不限制新闻消息的内容，例如可以包括新闻事件、热点话题、人物动态、产品资讯等中的至少一种；另外，也不限制新闻消息的实现格式，例如可以包括文本、图片、视频等中的至少一种。

另外，本申请实施例中的资源类别是指网络资源所属的类别。本申请实施例不限定网络资源的类型。在不同应用场景中，网络资源会有所不同，网络资源所属的类别也会有所不同。举例说明：

在电子商务领域，网络资源可以是卖家提供的各种商品、服务等，相应的，资源类别可以是网络资源所属的类目，例如女装、男装、鞋子、生活、学习、运动、户外、母婴等。值得说明的是，本申请实施例并不限制类目等级，也就是说，在本申请实施例中，资源类别可以包括各种等级的类目。

基于上述介绍，本实施例的业务处理方法具体包括：

首先，按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。

上述步骤中的抓取周期可以根据应用场景适应性设置，例如可以是一天、一周、三天、五天等。另外，考虑到网络平台上新闻消息的数量较多，并且这些新闻消息的价值会随着时间的增长而递减，所以在本实施例中，预先设定要求，具体抓取满足预设要求的新闻消息，这样可以降低新闻消息的数量，提高处理效率。所述预设要求可是热度大于指定热度阈值（这样可以获取热点新闻消息），或者是出现时间晚于指定时间（这样可以获取近期出现的新闻消息）。

例如，可以利用爬虫，将各大新闻网站（例如新浪网，其网址为 www.sina.com.cn，搜狐网，其网址为 www.sohu.com 等）上的热点新闻抓取下来。所谓热点新闻也就是热度比较高的新闻消息，例如具体可能是各大新闻网站上比较靠前的新闻消息，如头条消息等。优选地，这里的爬虫可采用 Jsoup 定向抓取技术，但不限于此。

另外，抓取到的新闻消息可以存储在数据存储系统，所述数据存储系统

可采用 mysql 关系型数据库实现，但不限于此。

在抓取到新闻消息之后，确定与抓取到的新闻消息相匹配的目标资源类别。

上述确定目标资源类别的一种可选实施方式包括：

计算新闻消息与资源类别库中各资源类别之间的相似度；

确定与新闻消息之间的相似度满足预设第一相似度条件的资源类别作为目标资源类别。

进一步，上述计算新闻消息与资源类别库中各资源类别之间的相似度的实施方式包括：

根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词；

对各资源类别分别进行分词处理，以获得各资源类别的关键词；

根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度。

更进一步，上述根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词的实施方式包括：

对新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得新闻消息的关键词。

更进一步，上述根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度的实施方式包括：

获取新闻消息的关键词的词向量和各资源类别的关键词的词向量；

根据新闻消息的关键词的词向量与各资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度。

在此说明，上述各步骤的具体实施方式可参见图 2a 所示实施例中相应步骤的描述，在此不再赘述。

在确定与抓取到的新闻消息相匹配的目标资源类别之后，对目标资源类

别下的网络资源进行业务处理。

值得说明的是，根据应用场景的不同，对目标资源类别下的网络资源进行业务处理的具体流程也会有所不同。下面以电子商务领域中的部分业务场景为例进行举例说明，对本领域技术人员来说，在下述举例说明的基础上，可以实现其他应用场景中对目标资源类别下的网络资源进行业务处理的流程。

在电子商务领域中，电商平台向用户推荐商品是一种较为常见的业务场景。由于热门新闻和资讯会影响商品价格和热度，所以可以采用本实施例提供的方法向用户推荐商品。具体的，按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；确定与新闻消息相匹配的目标资源类目；对目标资源类目下的商品进行推荐处理。

其中，上述对目标资源类目下的商品进行推荐处理包括但不限于以下处理：

确定目标资源类目下的商品是否具有推荐价值，以及在确定该商品具有推荐价值需要进行推荐时的推荐力度以及推荐方式等。

由上述可见，本实施例通过抓取新闻消息，确定与新闻消息相匹配的目标资源类别，进而对目标资源类别下的网络资源进行业务处理，提供一种基于新闻消息与资源类别之间的匹配关系的业务处理方法，充分发挥新闻消息对业务处理过程的影响，提高业务处理精度，同时丰富业务处理方式。

在此说明，本申请上述实施例提供的方法可以应用于电子商务领域，则网络资源可以是电商平台上的商品、网络资源所属的类别可以是商品类目，可以建立起热门新闻与商品类目之间的匹配关系，有利于让买家和卖家掌握一手的行业热点资讯，便于买家和卖家基于该匹配关系进行业务处理或决策。另外，本申请技术方案在实现上无需人工干预，能够自动捕捉热点新闻，达到智能化、自动化，效率较高。

需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中沒有详述的部分，可以参见其他实施例的相关描述。

图 4 为本申请又一实施例提供的业务处理装置的结构示意图。如图 4 所示，该装置包括：第一确定模块 41、获取模块 42 和业务模块 43。

第一确定模块 41，用于确定待处理网络资源所属的目标资源类别。

获取模块 42，用于获取与第一确定模块 41 所确定的目标资源类别相匹配的目标新闻消息。

业务模块 43，用于根据获取模块 42 获取的目标新闻消息，对待处理网络资源进行业务处理。

在一可选实施方式中，获取模块 42 具体可用于：

根据目标资源类别，查询预先建立的资源类别与新闻消息之间的匹配关系，以获取目标新闻消息。

在一可选实施方式中，如图 5 所示，该装置还包括：抓取模块 51、计算模块 52、第二确定模块 53 和建立模块 54。

抓取模块 51，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。

计算模块 52，用于计算新闻消息与资源类别库中各资源类别之间的相似度。

第二确定模块 53，用于确定与新闻消息之间的相似度满足预设第一相似度条件的资源类别。

建立模块 54，用于建立新闻消息和第二确定模块 53 所确定的资源类别之间的匹配关系。

进一步，如图 5 所示，计算模块 52 的一种实现结构包括：获取单元 521、分词单元 522 和计算单元 523。

获取单元 521，用于根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词。

分词单元 522，用于对各资源类别分别进行分词处理，以获得各资源类别的关键词。

计算单元 523, 用于根据获取单元 521 获取的新闻消息的关键词与分词单元 522 获得的各资源类别的关键词, 计算新闻消息与所述各资源类别之间的相似度。

进一步, 获取单元 521 具体用于:

对新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理, 以获得正文关键词、标题关键词和评论关键词中的至少一种;

将正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理, 以获得新闻消息的关键词。

进一步, 计算单元 523 具体用于:

获取新闻消息的关键词的词向量和各资源类别的关键词的词向量;

根据新闻消息的关键词的词向量与各资源类别的关键词的词向量, 计算新闻消息与各资源类别之间的相似度。

在一可选实施方式中, 获取模块 42 具体用于:

计算新闻语料库中各新闻消息与目标资源类别之间的相似度;

获取与目标资源类别之间的相似度满足预设第二相似度条件的新闻消息作为目标新闻消息。

进一步地, 获取模块 42 在计算新闻语料库中各新闻消息与目标资源类别之间的相似度时, 具体用于:

对目标资源类别进行分词处理, 以获得目标资源类别的关键词;

对于每个新闻消息, 根据新闻消息的正文、标题和评论信息中的至少一类信息, 获取新闻消息的关键词, 根据新闻消息的关键词与目标资源类别的关键词, 计算新闻消息与目标资源类别之间的相似度。

更进一步地, 获取模块 42 在根据新闻消息的关键词与目标资源类别的关键词, 计算新闻消息与目标资源类别之间的相似度时, 具体用于:

获取新闻消息的关键词的词向量和目标资源类别的关键词的词向量;

根据新闻消息的关键词的词向量与目标资源类别的关键词的词向量, 计算新闻消息和目标资源类别之间的相似度。

可选的, 待处理网络资源可以是商品, 目标资源类别可以是商品所属的

类目。相应的，新闻消息和资源类别之间的匹配关系具体为新闻消息与商品类目之间的匹配关系。

本实施例提供的业务处理模块，确定待处理网络资源所属的目标资源类别，获取与该目标资源类别相匹配的目标新闻消息，根据目标新闻消息，对待处理资源网络进行业务处理，实现了基于新闻消息与资源类别之间的匹配关系的业务处理，充分发挥新闻消息对业务处理过程的影响，提高业务处理精度，同时丰富业务处理方式。

图 6 为本申请又一实施例提供的数据处理装置的结构示意图。如图 6 所示，该装置包括：抓取模块 61、计算模块 62、确定模块 63 和建立模块 64。

抓取模块 61，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。

计算模块 62，用于计算新闻消息与资源类别库中各资源类别之间的相似度。

确定模块 63，用于确定与新闻消息之间的相似度满足预设第一相似度条件的资源类别。

建立模块 64，用于建立新闻消息和确定模块 63 所确定的资源类别之间的匹配关系。

在一可选实施方式中，如图 7 所示，计算模块 62 的一种实现结构包括：获取单元 621、分词单元 622 和计算单元 623。

获取单元 621，用于根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词。

分词单元 622，用于对各资源类别分别进行分词处理，以获得各资源类别的关键词。

计算单元 623，用于根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度。

进一步地，获取单元 621 具体用于：

对新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得新闻消息的关键词。

更进一步地，计算单元 623 具体用于：

获取新闻消息的关键词的词向量和各资源类别的关键词的词向量；

根据新闻消息的关键词的词向量与各资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度。

可选的，这里的资源类别可以是商品所属的类目。相应的，新闻消息和资源类别之间的匹配关系具体为新闻消息与商品类目之间的匹配关系。

本实施例提供的数据处理装置，通过抓取新闻消息，计算新闻消息与各资源类别之间的相似度，根据相似度确定与新闻消息相匹配的资源类别，建立新闻消息与所确定的资源类别之间的匹配关系，为后续基于网络资源的业务处理提供条件。

图 8 为本申请又一实施例提供的业务处理装置的结构示意图。如图 8 所示，该装置包括：抓取模块 81、确定模块 82 和业务模块 83。

抓取模块 81，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息。

确定模块 82，用于确定与上述新闻消息相匹配的目标资源类别。

业务模块 83，用于对上述目标资源类别下的网络资源进行业务处理。

在一可选实施方式中，确定模块 82 的一种实现结构包括：计算单元和确定单元。

计算单元，用于计算上述新闻消息与资源类别库中各资源类别之间的相似度；

确定单元，用于确定与上述新闻消息之间的相似度满足预设第一相似度条件的资源类别作为目标资源类别。

进一步，计算单元具体用于：

根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词；

对各资源类别分别进行分词处理，以获得各资源类别的关键词；

根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度。

更进一步，计算单元在根据新闻消息的正文、标题和评论信息中的至少一类信息，获取新闻消息的关键词时，具体用于：

对新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得新闻消息的关键词。

更进一步，计算单元在根据新闻消息的关键词与各资源类别的关键词，计算新闻消息与各资源类别之间的相似度时，具体用于：

获取新闻消息的关键词的词向量和各资源类别的关键词的词向量；

根据新闻消息的关键词的词向量与各资源类别的关键词的词向量，计算新闻消息与各资源类别之间的相似度。

本实施例提供的业务处理装置，通过抓取新闻消息，确定与新闻消息相匹配的目标资源类别，进而对目标资源类别下的网络资源进行业务处理，提供一种基于新闻消息与资源类别之间的匹配关系的业务处理方法，充分发挥新闻消息对业务处理过程的影响，提高业务处理精度，同时丰富业务处理方式。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）或处理器（processor）执行本申请各个实施例所述方法的部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

权 利 要 求 书

1、一种业务处理方法，其特征在于，包括：

确定待处理网络资源所属的目标资源类别；

获取与所述目标资源类别相匹配的目标新闻消息；

根据所述目标新闻消息，对所述待处理网络资源进行业务处理。

2、根据权利要求1所述的方法，其特征在于，所述获取与所述目标资源类别相匹配的目标新闻消息，包括：

根据所述目标资源类别，查询预先建立的资源类别与新闻消息之间的匹配关系，以获取所述目标新闻消息。

3、根据权利要求2所述的方法，其特征在于，所述建立资源类别与新闻消息之间的匹配关系，包括：

按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算所述新闻消息与资源类别库中各资源类别之间的相似度；

确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别；

建立所述新闻消息和所述确定的资源类别之间的匹配关系。

4、根据权利要求3所述的方法，其特征在于，所述计算新闻消息与资源类别库中各资源类别之间的相似度，包括：

根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词；

对所述各资源类别分别进行分词处理，以获得所述各资源类别的关键词；

根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度。

5、根据权利要求4所述的方法，其特征在于，所述根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词，包括：

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去

重处理，以获得所述新闻消息的关键词。

6、根据权利要求4或5所述的方法，其特征在于，所述根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度，包括：

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量，计算所述新闻消息与所述各资源类别之间的相似度。

7、根据权利要求1所述的方法，其特征在于，所述获取与所述目标资源类别相匹配的目标新闻消息，包括：

计算新闻语料库中各新闻消息与所述目标资源类别之间的相似度；

获取与所述目标资源类别之间的相似度满足预设第二相似度条件的新闻消息作为所述目标新闻消息。

8、根据权利要求7所述的方法，其特征在于，所述计算新闻语料库中各新闻消息与所述目标资源类别之间的相似度，包括：

对所述目标资源类别进行分词处理，以获得所述目标资源类别的关键词；

对于每个新闻消息，根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词，根据所述新闻消息的关键词与所述目标资源类别的关键词，计算所述新闻消息与所述目标资源类别之间的相似度。

9、根据权利要求8所述的方法，其特征在于，所述根据所述新闻消息的关键词与所述目标资源类别的关键词，计算所述新闻消息与所述目标资源类别之间的相似度，包括：

获取所述新闻消息的关键词的词向量和所述目标资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述目标资源类别的关键词的词向量，计算所述新闻消息和所述目标资源类别之间的相似度。

10、一种数据处理方法，其特征在于，包括：

按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算所述新闻消息与资源类别库中各资源类别之间的相似度；
确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别；
建立所述新闻消息和所述确定的资源类别之间的匹配关系。

11、根据权利要求 10 所述的方法，其特征在于，所述计算新闻消息与资源类别库中各资源类别之间的相似度，包括：

根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词；

对所述各资源类别分别进行分词处理，以获得所述各资源类别的关键词；

根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度。

12、根据权利要求 11 所述的方法，其特征在于，所述根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词，包括：

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得所述新闻消息的关键词。

13、根据权利要求 11 或 12 所述的方法，其特征在于，所述根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度，包括：

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量，计算所述新闻消息与所述各资源类别之间的相似度。

14、一种业务处理方法，其特征在于，包括：

按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

确定与所述新闻消息相匹配的目标资源类别；

对所述目标资源类别下的网络资源进行业务处理。

15、根据权利要求 14 所述的方法，其特征在于，所述确定与所述新闻消息

相匹配的目标资源类别，包括：

计算所述新闻消息与资源类别库中各资源类别之间的相似度；

确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别作为所述目标资源类别。

16、根据权利要求 15 所述的方法，其特征在于，所述计算新闻消息与资源类别库中各资源类别之间的相似度，包括：

根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词；

对所述各资源类别分别进行分词处理，以获得所述各资源类别的关键词；

根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度。

17、根据权利要求 16 所述的方法，其特征在于，所述根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词，包括：

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得所述新闻消息的关键词。

18、根据权利要求 16 或 17 所述的方法，其特征在于，所述根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度，包括：

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量，计算所述新闻消息与所述各资源类别之间的相似度。

19、一种业务处理装置，其特征在于，包括：

第一确定模块，用于确定待处理网络资源所属的目标资源类别；

获取模块，用于获取与所述目标资源类别相匹配的目标新闻消息；

业务模块，用于根据所述目标新闻消息，对所述待处理网络资源进行业务

处理。

20、根据权利要求 19 所述的装置，其特征在于，所述获取模块具体用于：

根据所述目标资源类别，查询预先建立的资源类别与新闻消息之间的匹配关系，以获取所述目标新闻消息。

21、根据权利要求 20 所述的装置，其特征在于，还包括：

抓取模块，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算模块，用于计算所述新闻消息与资源类别库中各资源类别之间的相似度；

第二确定模块，用于确定与所述新闻消息之间的相似度满足预设第一相似条件度的资源类别；

建立模块，用于建立所述新闻消息和所述确定的资源类别之间的匹配关系。

22、根据权利要求 21 所述的装置，其特征在于，所述计算模块包括：

获取单元，用于根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词；

分词单元，用于对所述各资源类别分别进行分词处理，以获得所述各资源类别的关键词；

计算单元，用于根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度。

23、根据权利要求 22 所述的装置，其特征在于，所述获取单元具体用于：

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得所述新闻消息的关键词。

24、根据权利要求 22 或 23 所述的装置，其特征在于，所述计算单元具体用于：

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量，计算所述新闻消息与所述各资源类别之间的相似度。

25、根据权利要求 19 所述的装置，其特征在于，所述获取模块具体用于：
计算新闻语料库中各新闻消息与所述目标资源类别之间的相似度；

获取与所述目标资源类别之间的相似度满足预设第二相似度条件的新闻消息作为所述目标新闻消息。

26、根据权利要求 25 所述的装置，其特征在于，所述获取模块具体用于：
对所述目标资源类别进行分词处理，以获得所述目标资源类别的关键词；

对于每个新闻消息，根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词，根据所述新闻消息的关键词与所述目标资源类别的关键词，计算所述新闻消息与所述目标资源类别之间的相似度。

27、根据权利要求 26 所述的装置，其特征在于，所述获取模块具体用于：
获取所述新闻消息的关键词的词向量和所述目标资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述目标资源类别的关键词的词向量，计算所述新闻消息和所述目标资源类别之间的相似度。

28、一种数据处理装置，其特征在于，包括：

抓取模块，用于按照预设抓取周期，从网络平台上抓取满足预设要求的新闻消息；

计算模块，用于计算所述新闻消息与资源类别库中各资源类别之间的相似度；

确定模块，用于确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别；

建立模块，用于建立所述新闻消息和所述确定的资源类别之间的匹配关系。

29、根据权利要求 28 所述的装置，其特征在于，所述计算模块包括：

获取单元，用于根据所述新闻消息的正文、标题和评论信息中的至少一类信息，获取所述新闻消息的关键词；

分词单元，用于对所述各资源类别分别进行分词处理，以获得所述各资源

类别的关键词;

计算单元, 用于根据所述新闻消息的关键词与所述各资源类别的关键词, 计算所述新闻消息与所述各资源类别之间的相似度。

30、根据权利要求 29 所述的装置, 其特征在于, 所述获取单元具体用于:

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理, 以获得正文关键词、标题关键词和评论关键词中的至少一种;

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理, 以获得所述新闻消息的关键词。

31、根据权利要求 28 或 29 所述的装置, 其特征在于, 所述计算单元具体用于:

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量;

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量, 计算所述新闻消息与所述各资源类别之间的相似度。

32、一种业务处理装置, 其特征在于, 包括:

抓取模块, 用于按照预设抓取周期, 从网络平台上抓取满足预设要求的新闻消息;

确定模块, 用于确定与所述新闻消息相匹配的目标资源类别;

业务模块, 用于对所述目标资源类别下的网络资源进行业务处理。

33、根据权利要求 32 所述的装置, 其特征在于, 所述确定模块包括:

计算单元, 用于计算所述新闻消息与资源类别库中各资源类别之间的相似度;

确定单元, 用于确定与所述新闻消息之间的相似度满足预设第一相似度条件的资源类别作为所述目标资源类别。

34、根据权利要求 33 所述的装置, 其特征在于, 所述计算单元具体用于:

根据所述新闻消息的正文、标题和评论信息中的至少一类信息, 获取所述新闻消息的关键词;

对所述各资源类别分别进行分词处理, 以获得所述各资源类别的关键词;

根据所述新闻消息的关键词与所述各资源类别的关键词，计算所述新闻消息与所述各资源类别之间的相似度。

35、根据权利要求 34 所述的装置，其特征在于，所述计算单元具体用于：

对所述新闻消息的正文、标题和评论消息中的至少一类信息进行关键词提取处理，以获得正文关键词、标题关键词和评论关键词中的至少一种；

将所述正文关键词、标题关键词和评论关键词中的至少一种进行合并和去重处理，以获得所述新闻消息的关键词。

36、根据权利要求 34 或 35 所述的装置，其特征在于，所述计算单元具体用于：

获取所述新闻消息的关键词的词向量和所述各资源类别的关键词的词向量；

根据所述新闻消息的关键词的词向量与所述各资源类别的关键词的词向量，计算所述新闻消息与所述各资源类别之间的相似度。

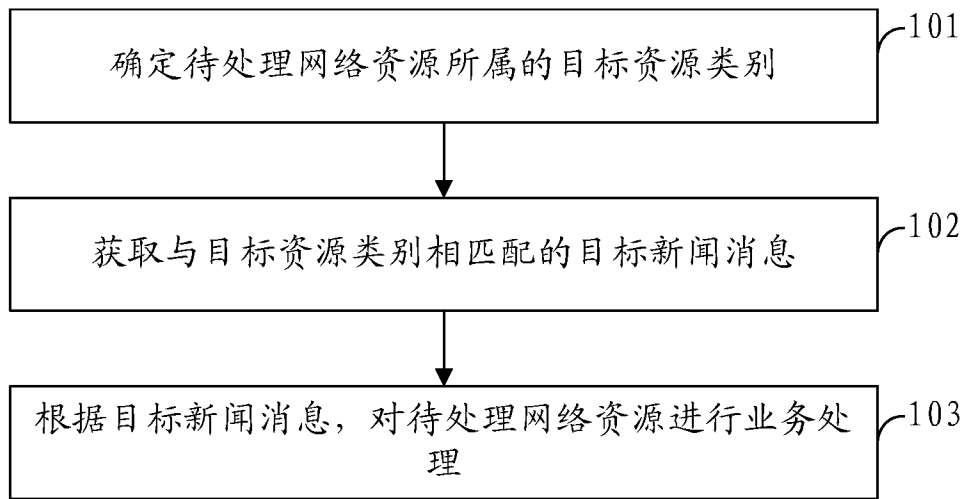


图 1

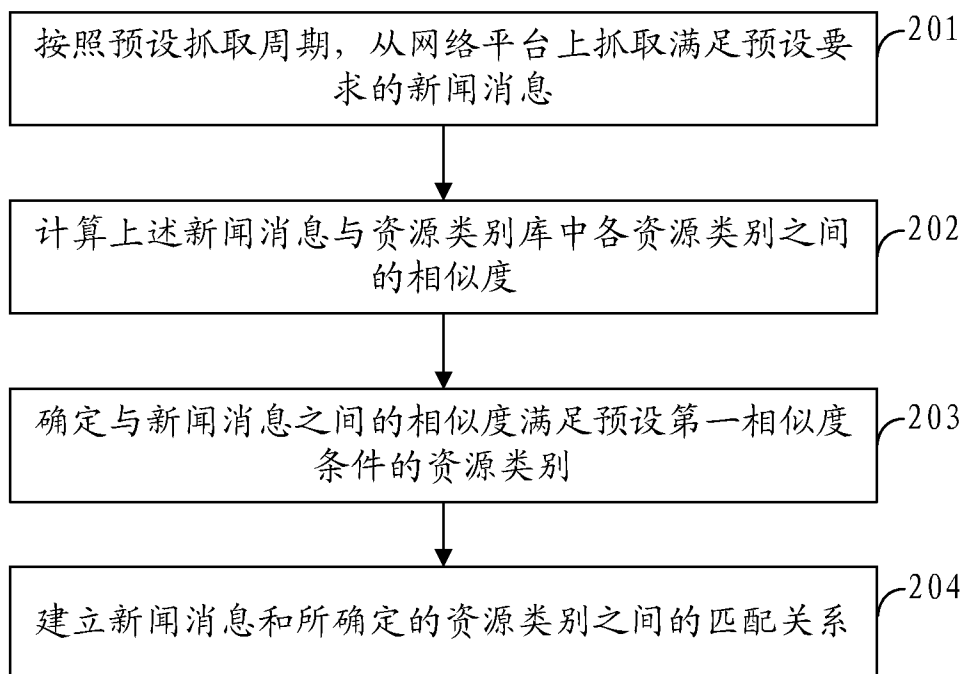


图 2a

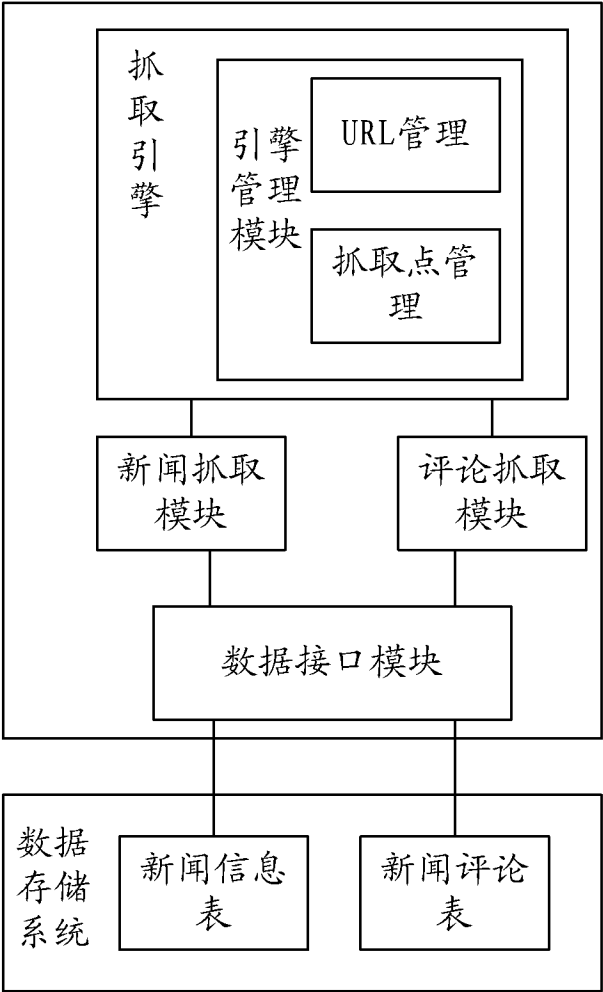


图 2b

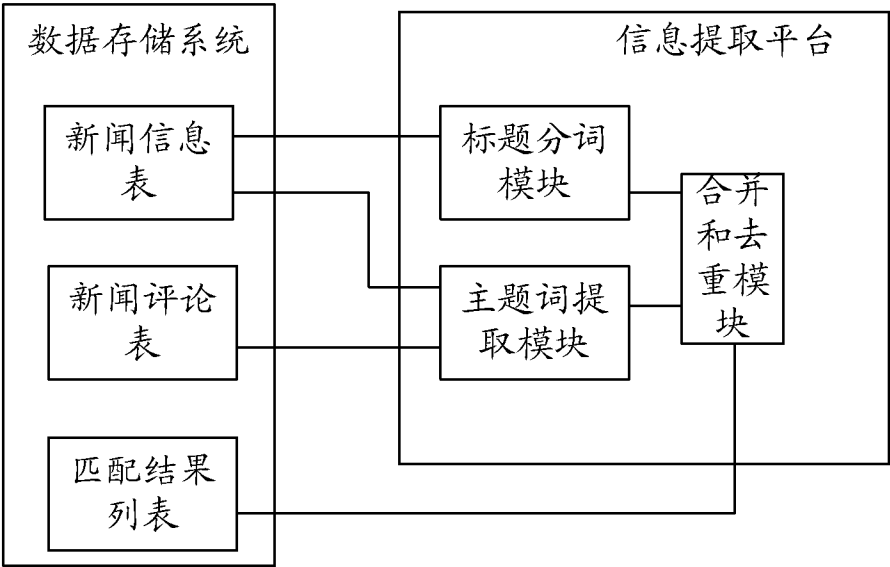


图 2c

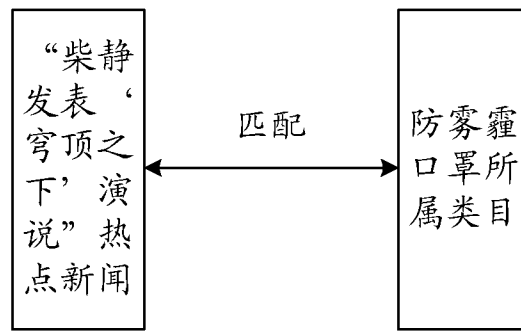


图 2d

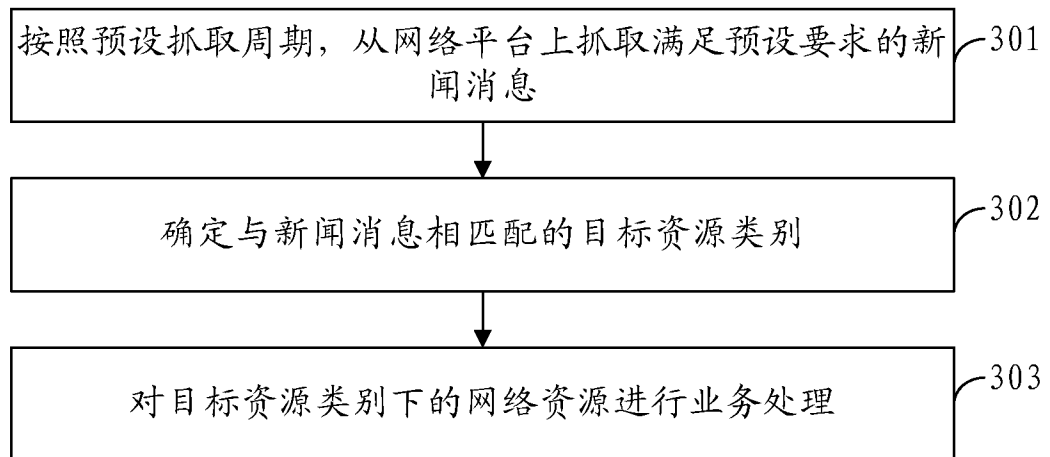


图 3

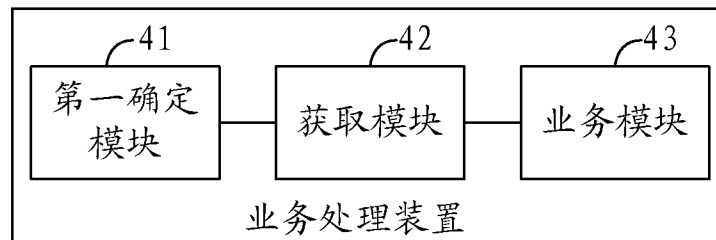


图 4

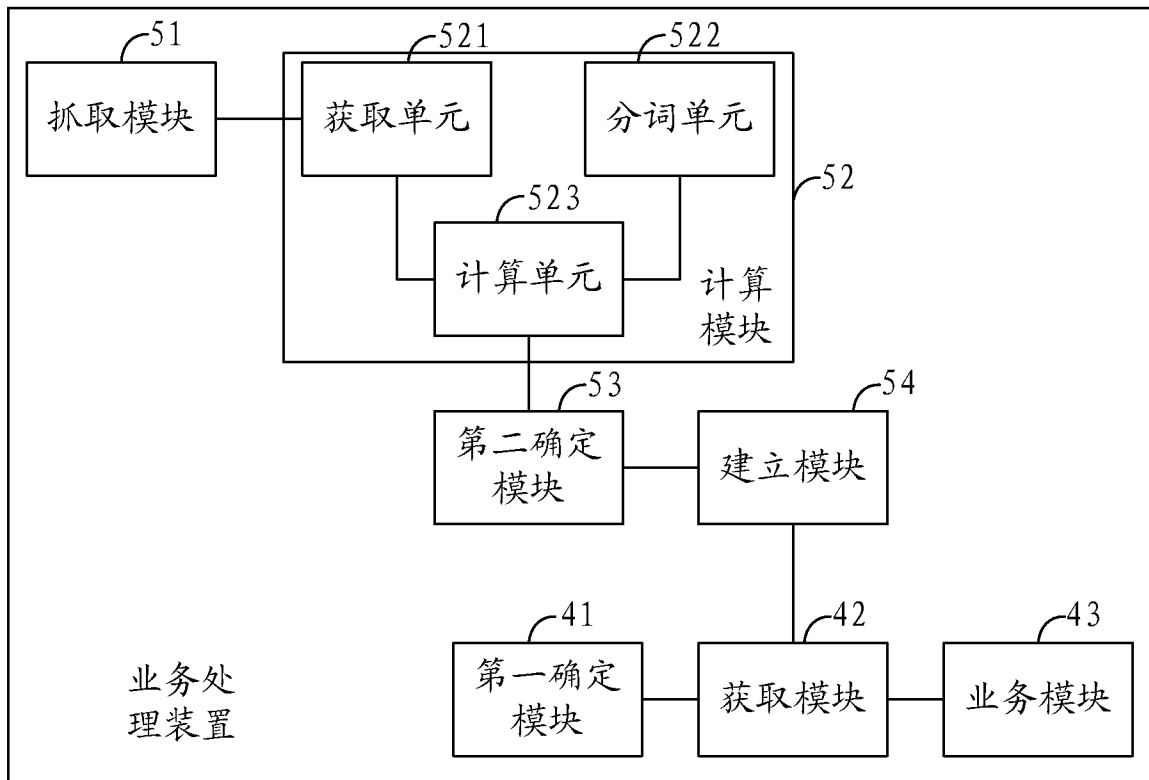


图 5

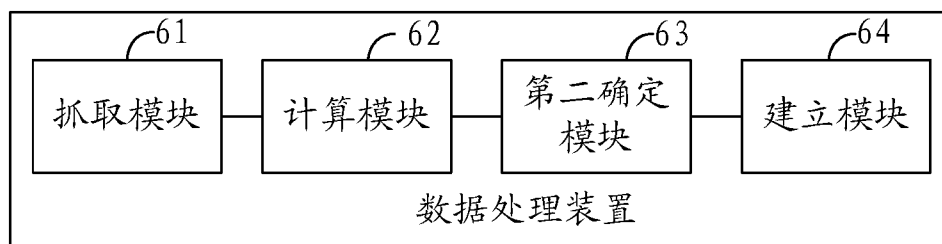


图 6

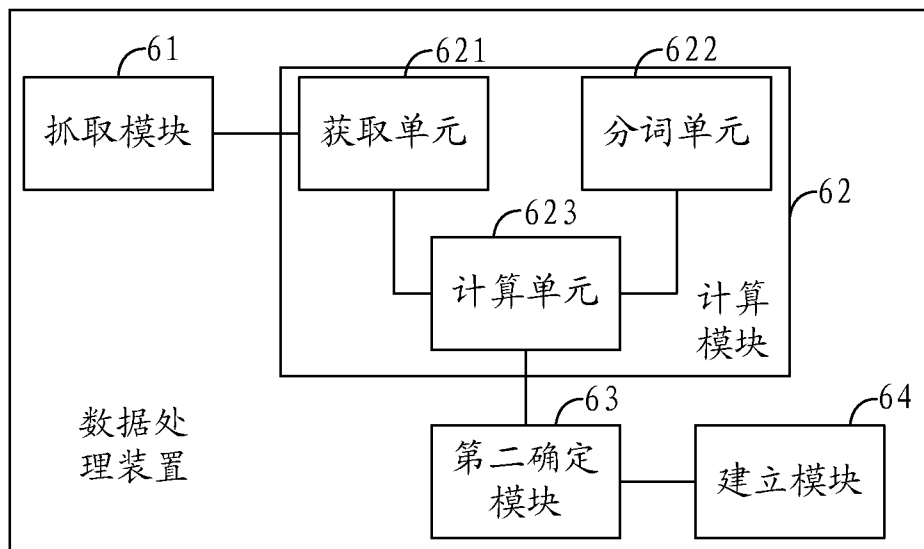


图 7

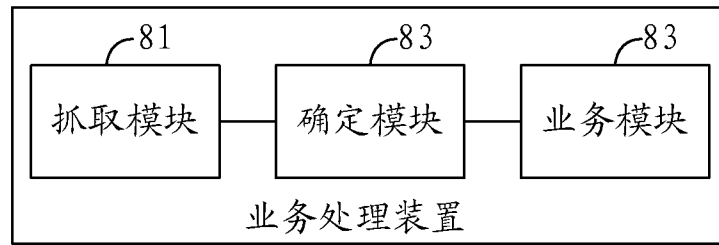


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/071409

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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 17/-

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)

CNKI, CNTXT, CNABS, DWPI, SIPOABS, USTXT: hot spot, news, hot, matching, similarity, influence

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103226554 A (XIZANG TONGXIN SECURITIES CO., LTD.), 31 July 2013 (31.07.2013), description, paragraphs 2-21	1, 10, 14-15, 19, 28, 32-33
A	CN 104115178 A (THOMSON REUTERS GLOBAL RESOURCES), 22 October 2014 (22.10.2014), the whole document	1-36
A	KR 20020015198 A (PAROLE SCIENCE INC.), 27 February 2002 (27.02.2002), the whole document	1-36

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 28 March 2017 (28.03.2017)	Date of mailing of the international search report 26 April 2017 (26.04.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer WANG, Yankun Telephone No.: (86-10) 62089282

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2017/071409

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103226554 A	31 July 2013	None	
CN 104115178 A	22 October 2014	WO 2013082527 A1	06 June 2013
		EP 2786335 A1	08 October 2014
		EP 2786335 A4	15 July 2015
		US 2013138577 A1	30 May 2013
KR 20020015198 A	27 February 2002	None	

国际检索报告

国际申请号

PCT/CN2017/071409

A. 主题的分类

G06F 17/30(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 17/-

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))

CNKI, CNTXT, CNABS, DWPI, SIPOABS, USTXT(新闻, 热点, 匹配, 相似, 影响, news, hot, matching, similarity, influence)

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 103226554 A (西藏同信证券有限责任公司) 2013年 7月 31日 (2013 - 07 - 31) 说明书第2-21段	1, 10, 14-15, 19, 28, 32-33
A	CN 104115178 A (汤姆森路透社全球资源公司) 2014年 10月 22日 (2014 - 10 - 22) 全文	1-36
A	KR 20020015198 A (PAROLE SCIENCE INC) 2002年 2月 27日 (2002 - 02 - 27) 全文	1-36

☐ 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

2017年 3月 28日

国际检索报告邮寄日期

2017年 4月 26日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

王艳坤

电话号码 (86-10)62089282

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/071409

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	103226554	A	2013年 7月 31日	无			
CN	104115178	A	2014年 10月 22日	WO	2013082527	A1	2013年 6月 6日
				EP	2786335	A1	2014年 10月 8日
				EP	2786335	A4	2015年 7月 15日
				US	2013138577	A1	2013年 5月 30日
KR	20020015198	A	2002年 2月 27日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)