

(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
2016年6月2日 (02.06.2016)



(10) 国际公布号
WO 2016/082575 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/086095
- (22) 国际申请日: 2015年8月5日 (05.08.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410710424.7 2014年11月27日 (27.11.2014) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司 (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 刘松 (LIU, Song); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100190 (CN)。 孙凯 (SUN, Kai); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100190 (CN)。 陶明远 (TAO, Mingyuan); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100190 (CN)。
- (74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路39号西金大厦6层, Beijing 100036 (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, 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, KN, KP, KR, 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)。

[见续页]

(54) Title: INFORMATION MINING METHOD AND APPARATUS, AND STORAGE MEDIUM

(54) 发明名称: 信息挖掘方法、装置和存储介质

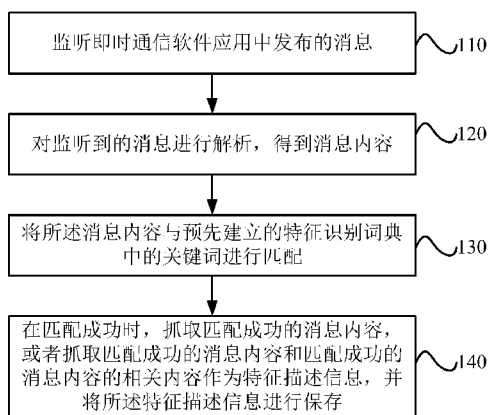


图1 / FIG. 1

110 MONITORING A MESSAGE ISSUED IN AN INSTANT COMMUNICATION SOFTWARE APPLICATION
120 PARSING THE MONITORED MESSAGE TO OBTAIN MESSAGE CONTENT
130 MATCHING THE MESSAGE CONTENT WITH A KEYWORD IN A PRE-ESTABLISHED FEATURE IDENTIFICATION DICTIONARY
140 WHEN THE MATCHING SUCCEEDS, CAPTURING THE SUCCESSFULLY MATCHED MESSAGE CONTENT OR CAPTURING THE SUCCESSFULLY MATCHED MESSAGE CONTENT AND RELATED CONTENT OF THE SUCCESSFULLY MATCHED MESSAGE CONTENT AS FEATURE DESCRIPTION INFORMATION, AND SAVING THE FEATURE DESCRIPTION INFORMATION

(57) Abstract: Provided are an information mining method and apparatus, and a storage medium. The method includes: monitoring a message issued in an instant communication software application (110); parsing the monitored message to obtain message content (120); matching the message content with a keyword in a pre-established feature identification dictionary (130); and when the matching succeeds, capturing the message content or the message content and related content of the message content as feature description information, and saving the feature description information (140). The message issued in the instant communication software application has high category definition and high information specialism, such that the feature description information of a specific object can be automatically captured by matching the parsed message content with the keyword in the feature identification dictionary and capturing the successfully matched message content or capturing the successfully matched message content and the related content of the message content, thereby reducing the labour costs, and improving the specialism and accuracy of the obtained feature description information of the specific object.

(57) 摘要: 提供一种信息挖掘方法、装置和存储介质。该方法包括: 监听即时通信软件应用中发布的消息(110); 对监听到的消息进行解析, 得到消息内容(120); 将消息内容与预先建立的特征识别词典中的关键词进行匹配(130); 在匹配成功时, 抓取所述消息内容, 或者所述消息内容和所述消息内容的相关内容作为特征描述信息, 并将所述特征描述信息进行保存(140)。由于即时通信软件应用中发布消息不仅类别清晰度高, 且信息专业性高, 因此通过将解析到的消息内容与特征识别词典中的关键词进行匹配, 并

抓取匹配成功的消息内容, 或抓取匹配成功的消息内容和该消息内容的相关内容, 可自动捕获特定对象的特征描述信息, 节省了人力成本, 并提升了得到的特定对象的特征描述信息的专业性和准确性。

WO 2016/082575 A1



本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

信息挖掘方法、装置和存储介质

本专利申请要求于 2014 年 11 月 27 日提交的、申请号为 201410710424.7、申请人为百度在线网络技术（北京）有限公司、发明名称为“信息挖掘方法和装置”的中国专利申请的优先权，该申请的全文以引用的方式并入本申请中。

技术领域

本发明实施例涉及信息技术领域，尤其涉及一种信息挖掘方法、装置和存储介质。

背景技术

现有技术中获取与产品或服务对象相关的信息，比如对产品的改进有帮助的产品缺陷描述信息时，通常是通过人工在相关领域的论坛或网页中进行抓取，效率低下且准确度不高。

发明内容

本发明实施例提供一种信息挖掘方法、装置和存储介质，以实现自动捕获特定对象的特征信息，节省人力成本，并提升捕获到的特定对象的特征信息的准确度。

第一方面，本发明实施例提供了一种信息挖掘方法，包括：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

第二方面，本发明实施例还提供了一种信息挖掘装置，包括：

消息监听模块，用于监听即时通信软件应用中发布的消息；

消息解析模块，用于对监听到的消息进行解析，得到消息内容；

匹配模块，用于将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

特征描述信息处理模块，用于在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

第三方面，本发明实施例还提供了一种非易失性计算机存储介质，所述计算机存储介质存储有一个或者多个模块，当所述一个或者多个模块被一个执行信息挖掘方法的设备执行时，使得所述设备执行如下操作：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

本发明实施例提供的信息挖掘方法、装置和存储介质，通过监听并解析即时通信软件应用中发布的消息，由于即时通信软件应用中发布消息不仅类别清晰度高，而且信息专业性高，因此通过将解析到的消息内容与预先建立的特征识别词典中的关键词进行匹配，并抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和该消息内容的相关内容，可以自动捕获特定对象的特征描述信息，节省了人力成本，并提升了得到的特定对象的特征描述信息的专业性和准确性，有利于根据所述特征描述信息对特定对象进行改进。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所

需使用的附图作简单地介绍，当然，以下描述中的附图仅仅是本发明的一些实施例，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，还可以对这些附图进行修改和替换。

图 1 为本发明实施例一提供的一种信息挖掘方法的流程图；

图 2 为本发明实施例二提供的一种信息挖掘方法的流程图；

图 3a 为本发明实施例三提供的一种信息挖掘方法的流程图；

图 3b 为本发明实施例三提供的另一种信息挖掘方法的流程图；

图 3c 为本发明实施例三提供的又一种信息挖掘方法的流程图；

图 4 为本发明实施例四提供的一种信息挖掘装置的结构示意图；

图 5 为本发明实施例六提供的一种执行信息挖掘方法的设备的硬件结构示意图。

具体实施方式

下面将结合附图对本发明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发明的一部分实施例，而不是全部的实施例，是为了阐述本发明的原理，而不是要将本发明限制于这些具体的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

实施例一

请参阅图 1，为本发明实施例一提供的一种信息挖掘方法的流程图。本发明实施例的方法可以由配置以硬件和/或软件实现的信息挖掘装置来执行，该实现装置典型的是配置于能够提供数据挖掘服务的服务器中。

该方法包括：操作 110~操作 140。

110、监听即时通信软件应用中发布的消息。

通常，每个企业内部都有跟该企业产品或部门相关的即时通信软件应用，以方便该企业内负责各产品研发群体或负责运营维护的群体发布消息。

例如，百度公司推出的百度 Hi 是一款集文字消息、语音视频通话和文件传

输等功能的即时通讯软件应用，在百度 Hi 中可以建立跟产品“百度地图”或产品“百度翻译”等对应的群组，以方便百度公司内负责各产品研发或负责运营维护的工作人员发布消息。

其中，发布消息的方式有多种，可以以文字形式发布，也可以以语音、视频或图片等其他形式发布，本实施例对此不进行限制，只要得到即时通信软件应用支持即可。

本操作具体可以是监听即时通信软件应用中与企业产品相关的群组或与企业部门相关的群组中发布的文字消息。若为其他形式发布的消息，需要通过相关识别技术（例如，语音识别技术或图片识别技术）识别其中的内容，得到对应的文字消息。

120、对监听到的消息进行解析，得到消息内容。

本操作中，具体是根据即时通信软件应用的通信协议，对监听到的消息进行翻译，正确还原出与监听到的消息对应的原始数据，也即还原出可阅读的字符串。

130、将所述消息内容与预先建立的特征识别词典中的关键词进行匹配。

本操作具体是利用关键词匹配技术，根据预先建立的特征识别词典，确定所述消息内容中是否包含所述特征识别词典中的关键词。

需要说明的是，企业内各对象对应的群体发布消息不同，解析到的消息内容不同。群体具有类别清晰度高、信息专业性高和语言特征明显的特点（例如每个群组包含的群成员都是一种类别或者做同一产品的人群，群成员都具相同或相似的专业背景），因此不同群体发布的消息能够反映企业对象信息。

其中，对象可以是具体的各个产品，也可以是企业管理等宏观对象。

例如，“百度地图”产品对应的群组是百度公司负责“百度地图”研发或运营维护的群体，该群组中群成员发布的消息通常包含有该产品的优缺点信息、或该产品的后续改进信息。

又如，“百度浏览器”产品对应的调试群组中群成员发布的消息通常包含有该产品调试过程中出现的 bug 或者疑似问题。

因此，可以对企业不同的对象对应的群组建立相应的特征识别词典，从而

得到不同对象（例如不同的产品，或者企业管理）对应的特征描述信息（例如不同产品的优缺点信息，或企业管理存在的问题）；对企业同一对象的不同群组，优选是建立相应的特征识别词典，从而得到与同一对象有关的不同层面的特征描述信息。

例如，对“百度地图”产品中的研发群体建立与研发有关的特征识别词典，该词典中的关键词可以包括“研发”、“进展”、“趋势”、“成本”和“对手”等；对“百度地图”产品中的调试群体建立与调试有关的特征识别词典，该词典中的关键词可以包括“调试错误”、“调试周期”、“bug”、“漏洞”和“缺陷”等；对“百度地图”产品中的发布群体建立与发布有关的特征识别词典，该词典中的关键词可以包括“发布”、“发布会”、“发布行程”和“发布日期”等。

140、在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

本操作中，可以有两种实施方式，一种是在匹配成功时，抓取匹配成功的消息内容作为特征描述信息，并将所述特征描述信息进行保存；另一种是在匹配成功时，抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

其中，优选是抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存，相比于只抓取匹配成功的消息内容，该优选方式有利于得到对象的完整的特征描述信息。

可以设定抓取时间间隔和/或抓取条数，以抓取匹配成功的消息内容的相关内容，例如将抓取时间间隔设定为 15s，将抓取条数设定为 5。

进一步地，所述匹配成功的消息内容的相关内容可以包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

示例 1

以对象为“百度浏览器”产品为例进行说明。该产品的某个群组发布的消

息中包含大量有关该产品的评价和问题讨论，例如：该产品的一个设计人员在开发群组中发布消息“登录百度浏览器时，登录权限有问题”，接着该产品的另一个设计人员在该开发群组中发布消息“确实，原因是A”；通过匹配操作后，“登录百度浏览器时，登录权限有问题”这条发布消息与所述特征识别词典中的关键词“问题”匹配成功，通过抓取消息内容“登录百度浏览器时，登录权限有问题”，可以得到该产品的缺陷对应的特征描述信息，并通过抓取该消息内容的上下文消息“确实，原因是A”，可以得到该产品中该缺陷的产生原因对应的特征描述信息，从而丰富了该产品的特征描述信息。

需要说明的是，上述以抓取产品的缺陷对应的特征描述信息和该缺陷的产生原因对应的特征描述信息为例进行说明，除了抓取该缺陷的产生原因对应的特征描述信息之外，还可以抓取该缺陷对应的解决方案等其他特征描述信息，作为产品缺陷的完整信息，并进行格式化（例如[产品名称、缺陷内容、产生原因]）存储，本实施例对此不进行限制。

示例 2

在与发布所述消息内容的用户建立会话后，并采取启发式的提问向所述用户发送消息内容补充请求，以请求补充产品缺陷的完整描述，此时可以基于会话（session）进行抓取，即针对缺陷描述维度较多（例如缺陷类型、缺陷产生原因等）的情况，设置一个较长的抓取时间（如一分钟），在此时间内，抓取所述用户返回的补充内容。如果在此时间内仍没有补充描述，则只记录基本信息，或由于必要信息不全而返回失败。

本实施例的技术方案，通过监听并解析即时通信软件应用中发布的消息，由于即时通信软件应用中发布消息不仅类别清晰度高，而且信息专业性高，因此通过将解析到的消息内容与预先建立的特征识别词典中的关键词进行匹配，并抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和该消息内容的相关内容，可以自动捕获特定对象的特征描述信息，节省了人力成本，并提升了得到的特定对象的特征描述信息的专业性和准确性，有利于根据所述特征描述信息对特定对象进行改进。

在本实施例中，建立所述特征识别词典，具体可以包括：

接收人工配置的特征识别词典中的关键词；或者，

在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别词典中。

换言之，可以人工配置特征识别词典中的各关键词，例如，在特征识别词典中配置“问题”、“缺陷”或“改进”等关键词。

也可以人工收录一些典型语句，并根据聊天历史记录中典型语句的上下文共现关系，从而将满足一定共现频率的典型语句中的表达特征的词作为关键词，并添加至特征识别词典中；或者挖掘出表达特征的语义模板。

例如，在百度 Hi 的“百度浏览器”产品的研发群组中，一个人说“检索式=xxx，配图错误啊，谁谁看看”，另一个人回答“没错，是个问题，已记录缺陷”，如果群消息里多次出现“配图错误”和“已记录缺陷”这两句话配对时，就认为配对的这两句话存在共现关系，表明这个是需要记录的缺陷，基于此，可以挖掘出表达缺陷的语义模板“[任意词]配图错误”。

本实施例提供的信息挖掘方法，可以应用于多种场景，例如，根据建立的产品缺陷对应的特征识别词典，得到对象为产品的缺陷描述信息；又如，根据建立的产品调试对应的特征识别词典，得到对象为产品的调试问题描述信息；再如，根据建立的企业管理对应的特征识别词典，得到对象为企业管理事件的管理意见征集等描述信息，本实施例对此不进行限制。

具体地，当用于捕获对象为产品的缺陷描述信息时，所述特征识别词典中的关键词包含反映产品缺陷的关键词，所述特征描述信息为描述产品缺陷的信息。本实施方式提供了从产品挖掘，到缺陷相关内容抓取，最终保存到指定空间的全自动化的实现方式，可以覆盖企业所有产品线的重要产品群。

实施例二

请参阅图 2，为本发明实施例二提供的一种信息挖掘方法的流程图。本实施例在上述实施例的基础上，提供了在监听即时通信软件应用中发布的消息之前的优选方案。该优选方法包括：操作 210~操作 220。

210、在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务器建立连接。

例如，获取与即时通信软件应用“百度Hi”对应的服务器的访问权限，并与该服务器建立连接。

220、向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求。

例如，向即时通信软件应用“百度Hi”对应的服务器发送群组账号“百度浏览器-研发群”的加入请求，从而使得新加入的群成员能够在该群组中发布与产品“百度浏览器”有关的消息。

又如，向即时通信软件应用“百度Hi”对应的服务器发送个人用户账号加入请求，新加入的个人账号可以与已经加入该应用的其他个人账号就同一产品聊天，形成发布的消息；新加入的个人账号可以申请加入已经加入该应用的群组账号，从而使得新加入的群成员在该群组中发布消息。

本实施例的技术方案，在监听即时通信软件应用中发布的消息之前，通过与即时通信软件应用对应的服务器建立连接，并交互账号加入请求，从而使得加入该即时通信软件应用中的账号能够在该应用中发布消息。

需要说明的是，在向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求之后，监听即时通信软件应用中发布的消息，具体可以包括：在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

实施例三

请参阅图3a，为本发明实施例三提供的一种信息挖掘方法的流程图。本实施例在上述各实施例的基础上，提供了在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前的优选方案。

该优选方法包括：操作310~操作360。

310、监听即时通信软件应用中发布的消息。

320、对监听到的消息进行解析，得到消息内容。

330、将所述消息内容与预先建立的特征识别词典中的关键词进行匹配。

340、在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息。

350、将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别。

如前所述，本发明实施例提供的信息挖掘方法，可以应用于多种场景，因此可以根据实际应用需求，建立包含有多种应用需求的类别识别词典。

类别识别词典中的关键词可以人工配置。所述类别识别词典中的关键词可以包括：百度地图研发缺陷、百度浏览器调试缺陷和百度翻译研发改进等，本实施例对此不进行限制。

360、将确定的类别与所述特征描述信息进行关联保存。

本实施例的技术方案，通过监听并解析即时通信软件应用中发布的消息，由于即时通信软件应用中发布消息不仅类别清晰度高，而且信息专业性高，因此通过将解析到的消息内容与预先建立的特征识别词典中的关键词进行匹配，并抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和该消息内容的相关内容，可以自动捕获特定对象的特征描述信息，节省了人力成本，并提升了得到的特定对象的特征描述信息的专业性和准确性，有利于根据所述特征描述信息对特定对象进行改进；在抓取到对象的特征描述信息之后，通过确定所述特征描述信息对应的类别，并将确定的类别与所述特征描述信息进行关联保存，有利于绑定类别对应的负责群体，从而能够使相应负责群体根据特定对象的专业特征描述信息，及时获知对象的有价值的反馈。

需要说明的是，操作 350 只是确定特征描述信息对应的类别的实施方式中的其中一种，确定特征描述信息对应的类别还可以是：通过自然语言处理（Natural Language Processing NLP）模型确定所述特征描述信息对应的类别（如图 3b 所示的操作 351）。

具体可以采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

其中，语义相似度利用了自然语言处理云后台训练的监督方法训练模型来分析两段文本的相似度。值越大越相似。语义相似度的网络化提供了计算相似度的功能。比如输入“笔记本电脑”，“笔记本”的语义相似度为 2.08478。

其中，点击相似度可以在语义相似度无法达到阈值(如 1.8)的情况下使用，分析两段文本的点击相似度(比如检索式和检索结果中的标题)，使用训练的 embedding 向量计算 cosine 相似度值，取值范围[-1,1]，值越大点击相似度越强。比如输入“百度你好”和“周鸿祎你好”两者的点击相似度是-0.121407，输入“百度你好”和“李彦宏你好”两者的点击相似度是 0.218664；后者点击相似度比前者高。

实际使用中优先将特征描述信息与预设的多个类别分别进行语义相似度判断，返回语义相似度达到阈值且最高的类别，如果特征描述信息与预设类别的语义相似度未达到阈值，则继续将特征描述信息与该预设类别进行点击相似度判断，如果点击相似度达到阈值则返回相应类别，如果点击相似度未达到阈值，则返回默认类别(如：其他)。阈值会根据历史数据不断拟合，以保持更高的准确度。

还需要说明的是，确定特征描述信息对应的类别还可以是：采用预先根据已标注类别信息的特征描述文本训练出的概率模型确定所述特征描述信息对应的类别，所述概率模型的输入为特征描述文本，输出为属于设定类别的概率值(如图 3c 所示的操作 352)。具体的，根据已标注类别信息的特征描述文本预先训练出概率模型，将所述特征描述信息输入该概率模型，得到该概率模型输出的所述特征描述信息所对应的类别 A 及对应该类别 A 的概率值，若该概率值满足一定阈值，则确定所述特征描述信息对应的类别为类别 A。例如可以通过聊天记录中的人工分类标注和对应的描述文本，训练出 $P(\text{类型}|\text{特征描述信息})$ 的概率模型，训练方法可根据系统的业务领域特点灵活选择，典型的如朴素贝叶斯方法。在应用中，若用户问题描述属于某一问题分类的概率满足一定阈值，即可认为属于该分类。

在本实施例的基础上，在确定所述特征描述信息对应的类别之后，还可以包括下述操作：

根据所述类别确定所述特征描述信息的接收方的信息；

根据所述接收方的信息将所述特征描述信息发送给所述接收方。

其中，所述接收方的信息可以为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

本实施方式，提供了在抓取到对象的特征描述信息、并确定所述特征描述信息对应的类别之后，使接收方获知对象的特征描述信息的实现方式，将接收方作为类别对应的负责群体，并与该负责群体交互对象的专业特征描述信息，从而能够使相应负责群体根据特定对象的专业特征描述信息，及时获知对象的有价值的反馈。

实施例四

请参阅图 4，为本发明实施例四提供的一种信息挖掘装置的结构示意图。该装置包括：消息监听模块 410、消息解析模块 420、匹配模块 430 和特征描述信息处理模块 440。

其中，消息监听模块 410 用于监听即时通信软件应用中发布的消息；消息解析模块 420 用于对监听到的消息进行解析，得到消息内容；匹配模块 430 用于将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；特征描述信息处理模块 440 用于在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

本实施例的技术方案，通过监听并解析即时通信软件应用中发布的消息，由于即时通信软件应用中发布消息不仅类别清晰度高，而且信息专业性高，因此通过将解析到的消息内容与预先建立的特征识别词典中的关键词进行匹配，并抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和该消息内容的相关内容，可以自动捕获特定对象的特征描述信息，节省了人力成本，并提升了得到的特定对象的特征描述信息的专业性和准确性，有利于根据所述特征描述信息对特定对象进行改进。

在上述方案中，所述装置还可以包括：连接建立模块和请求发送模块。

其中，连接建立模块用于在监听即时通信软件应用中发布的消息之前，在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务器建立连接；请求发送模块用于向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求；所述消息监听模块 410 具体用于：在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

在上述方案中，所述装置还可以包括特征识别词典建立模块，用于接收人工配置的特征识别词典中的关键词；或者，

用于在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别词典中。

在上述方案中，所述装置还可以包括：第一类别确定模块、或第二类别确定模块、或第三类别确定模块。

其中，第一类别确定模块用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别；第二类别确定模块用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，通过自然语言处理（NLP）模型确定所述特征描述信息对应的类别；第三类别确定模块用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，采用预先根据已标注类别信息的特征描述文本训练出的概率模型确定所述特征描述信息对应的类别；所述特征描述信息处理模块 440 具体可用于：将确定的类别与所述特征描述信息进行关联保存。

其中，所述第二类别确定模块具体可用于：采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

进一步地，所述装置还可以包括：接收方信息确定模块和特征描述信息发送模块。

其中，接收方信息确定模块用于在确定所述特征描述信息对应的类别之后，根据所述类别确定所述特征描述信息的接收方的信息；特征描述信息发送模块用于根据所述接收方的信息将所述特征描述信息发送给所述接收方。

其中，所述接收方的信息可以为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

所述匹配成功的消息内容的相关内容可以包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

在上述方案中，所述特征识别词典中的关键词可以包含反映产品缺陷的关键词，相应地，所述特征描述信息可以为描述产品缺陷的信息。

本发明实施例提供的信息挖掘装置可执行本发明任意实施例所提供的信息挖掘方法，具备执行方法相应的功能模块和有益效果。

实施例五

本实施例提供一种非易失性计算机存储介质，所述计算机存储介质存储有一个或者多个模块，当所述一个或者多个模块被一个执行信息挖掘方法的设备执行时，使得所述设备执行如下操作：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

上述存储介质中存储的模块被所述设备执行时，在监听即时通信软件应用中发布的消息之前，还可包括：

在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务

器建立连接；

向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求；

所述监听即时通信软件应用中发布的消息，可具体包括：

在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

上述存储介质中存储的模块被所述设备执行时，建立所述特征识别词典，可具体包括：

接收人工配置的特征识别词典中的关键词；或者，

在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别词典中。

上述存储介质中存储的模块被所述设备执行时，在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，还可包括：

将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别；或，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别；或，采用预先根据已标注类别信息的特征描述文本训练出的概率模型确定所述特征描述信息对应的类别；

将所述特征描述信息进行保存可包括：将确定的类别与所述特征描述信息进行关联保存。

上述存储介质中存储的模块被所述设备执行时，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别，可具体包括：

采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

上述存储介质中存储的模块被所述设备执行时，在确定所述特征描述信息对应的类别之后，还可包括：

根据所述类别确定所述特征描述信息的接收方的信息；

根据所述接收方的信息将所述特征描述信息发送给所述接收方。

上述存储介质中存储的模块被所述设备执行时，所述接收方的信息可以为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

上述存储介质中存储的模块被所述设备执行时，所述匹配成功的消息内容的相关内容可以包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

上述存储介质中存储的模块被所述设备执行时，所述特征识别词典中的关键词可包含反映产品缺陷的关键词，所述特征描述信息为描述产品缺陷的信息。

实施例六

请参阅图 5，为本发明实施例六提供的一种执行信息挖掘方法的设备的硬件结构示意图。

该设备包括：

一个或者多个处理器 510，图 5 中以一个处理器 510 为例；

存储器 520；以及一个或者多个模块。

所述设备还可以包括：输入装置 530 和输出装置 540。所述设备中的处理器 510、存储器 520、输入装置 530 和输出装置 540 可以通过总线或其他方式连接，图 5 中通过总线连接为例。

存储器 520 作为一种计算机可读存储介质，可用于存储软件程序、计算机可执行程序以及模块，如本发明实施例中的信息挖掘方法对应的程序指令/模块（例如，附图 4 所示的信息挖掘装置中的消息监听模块 410，消息解析模块 420、匹配模块 430 和特征描述信息处理模块 440）。处理器 510 通过运行存储在存储器 520 中的软件程序、指令以及模块，从而执行服务器的各种功能应用以及数据处理，即实现上述方法实施例中的信息挖掘方法。

存储器 520 可包括存储程序区和存储数据区，其中，存储程序区可存储操

作系统、至少一个功能所需的应用程序；存储数据区可存储根据终端设备的使用所创建的数据等。此外，存储器 520 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中，存储器 520 可进一步包括相对于处理器 510 远程设置的存储器，这些远程存储器可以通过网络连接至终端设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

输入装置 530 可用于接收输入的数字或字符信息，以及产生与终端的用户设置以及功能控制有关的键信号输入。输出装置 540 可包括显示屏等显示设备。

所述一个或者多个模块存储在所述存储器 520 中，当被所述一个或者多个处理器 510 执行时，执行如下操作：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

进一步地，在监听即时通信软件应用中发布的消息之前，还可包括：

在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务器建立连接；

向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求；

所述监听即时通信软件应用中发布的消息，具体包括：

在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

进一步地，建立所述特征识别词典，可具体包括：

接收人工配置的特征识别词典中的关键词；或者，

在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别

词典中。

进一步地，在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，还可包括：

将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别；或，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别；或，采用预先根据已标注类别信息的特征描述文本训练出的概率模型确定所述特征描述信息对应的类别；

将所述特征描述信息进行保存可包括：将确定的类别与所述特征描述信息进行关联保存。

进一步地，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别，可具体包括：

采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

进一步地，在确定所述特征描述信息对应的类别之后，还可包括：

根据所述类别确定所述特征描述信息的接收方的信息；

根据所述接收方的信息将所述特征描述信息发送给所述接收方。

进一步地，所述接收方的信息可以为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

进一步地，所述匹配成功的消息内容的相关内容可以包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

进一步地，所述特征识别词典中的关键词可包含反映产品缺陷的关键词，所述特征描述信息为描述产品缺陷的信息。

通过以上关于实施方式的描述，所属领域的技术人员可以清楚地了解到，本发明可借助软件及必需的通用硬件来实现，当然也可以通过硬件实现，但很多情况下前者是更佳的实施方式。基于这样的理解，本发明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

软件产品可以存储在计算机可读存储介质中，如计算机的软盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、闪存（FLASH）、硬盘或光盘等，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，或者网络设备等等）执行本发明各个实施例所述的方法。

值得注意的是，上述信息挖掘装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限制本发明的保护范围。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明披露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以所述权利要求的保护范围为准。

权 利 要 求 书

1、一种信息挖掘方法，其特征在于，包括：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述信息进行保存。

2、如权利要求 1 所述的方法，其特征在于，在监听即时通信软件应用中发布的消息之前，还包括：

在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务器建立连接；

向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求；

所述监听即时通信软件应用中发布的消息，具体包括：

在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

3、如权利要求 1 或 2 所述的方法，其特征在于，建立所述特征识别词典，具体包括：

接收人工配置的特征识别词典中的关键词；或者，

在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别词典中。

4、如权利要求 1-3 任一所述的方法，其特征在于，在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，还包括：

将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别；或，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别；或，采用预先根据已标注类别信息的特

征描述文本训练出的概率模型确定所述特征描述信息对应的类别；

将所述特征描述信息进行保存包括：将确定的类别与所述特征描述信息进行关联保存。

5、如权利要求4所述的方法，其特征在于，通过自然语言处理NLP模型确定所述特征描述信息对应的类别，具体包括：

采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

6、如权利要求4或5所述的方法，其特征在于，在确定所述特征描述信息对应的类别之后，还包括：

根据所述类别确定所述特征描述信息的接收方的信息；

根据所述接收方的信息将所述特征描述信息发送给所述接收方。

7、如权利要求6所述的方法，其特征在于，所述接收方的信息为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

8、如权利要求1-7任一所述的方法，其特征在于，所述匹配成功的消息内容的相关内容包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

9、如权利要求1-8中任一所述的方法，其特征在于，所述特征识别词典中的关键词包含反映产品缺陷的关键词，所述特征描述信息为描述产品缺陷的信息。

10、一种信息挖掘装置，其特征在于，包括：

消息监听模块，用于监听即时通信软件应用中发布的消息；

消息解析模块，用于对监听到的消息进行解析，得到消息内容；

匹配模块，用于将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

特征描述信息处理模块，用于在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征

描述信息，并将所述特征描述信息进行保存。

11、如权利要求 10 所述的装置，其特征在于，所述装置还包括：

连接建立模块，用于在监听即时通信软件应用中发布的消息之前，在获取与所述即时通信软件应用对应的服务器的访问权限后，与所述服务器建立连接；

请求发送模块，用于向所述服务器发送对所述即时通信软件应用中的群组账号或个人用户账号的加入请求；

所述消息监听模块具体用于：在接收到所述服务器返回的同意加入的响应消息后，监听加入的群组中的用户或加入的个人用户发布的消息。

12、如权利要求 10 或 11 所述的装置，其特征在于，所述装置还包括特征识别词典建立模块，用于接收人工配置的特征识别词典中的关键词；或者，

用于在所述即时通信软件的聊天历史记录中查找人工收录的典型语句，根据该典型语句的上下文共现关系，挖掘出表达相应特征的关键词并添加在特征识别词典中。

13、如权利要求 10-12 任一所述的装置，其特征在于，所述装置还包括：

第一类别确定模块，用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，将所述特征描述信息与预先建立的类别识别词典中的关键词进行匹配，根据匹配结果确定所述特征描述信息对应的类别；或

第二类别确定模块，用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，通过自然语言处理 NLP 模型确定所述特征描述信息对应的类别；或

第三类别确定模块，用于在抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息之后、将所述特征描述信息进行保存之前，采用预先根据已标注类别信息的特征描述文本训练出的概率模型确定所述特征描述信息对应的类别；

所述特征描述信息处理模块具体用于：将确定的类别与所述特征描述信息进行关联保存。

14、如权利要求 13 所述的装置，其特征在于，所述第二类别确定模块具体用于：采用语义相似度算法模型和/或点击相似度算法模型，确定所述的特征描述信息对应的类别。

15、如权利要求 13 或 14 所述的装置，其特征在于，所述装置还包括：

接收方信息确定模块，用于在确定所述特征描述信息对应的类别之后，根据所述类别确定所述特征描述信息的接收方的信息；

特征描述信息发送模块，用于根据所述接收方的信息将所述特征描述信息发送给所述接收方。

16、如权利要求 15 所述的装置，其特征在于，所述接收方的信息为设定网站的地址、设定接收用户的短信号码、邮箱地址或设定接收用户的即时通信软件账号。

17、如权利要求 10-16 任一所述的装置，其特征在于，所述匹配成功的消息内容的相关内容包括：所述匹配成功的消息内容的上下文消息；和/或，在与发布所述消息内容的用户建立会话并向所述用户发送消息内容补充请求后，所述用户返回的补充内容。

18、如权利要求 10-17 中任一所述的装置，其特征在于，所述特征识别词典中的关键词包含反映产品缺陷的关键词，所述特征描述信息为描述产品缺陷的信息。

19、一种非易失性计算机存储介质，所述计算机存储介质存储有一个或者多个模块，其特征在于，当所述一个或者多个模块被一个执行信息挖掘方法的设备执行时，使得所述设备执行如下操作：

监听即时通信软件应用中发布的消息；

对监听到的消息进行解析，得到消息内容；

将所述消息内容与预先建立的特征识别词典中的关键词进行匹配；

在匹配成功时，抓取匹配成功的消息内容，或者抓取匹配成功的消息内容和所述匹配成功的消息内容的相关内容作为特征描述信息，并将所述特征描述

信息进行保存。

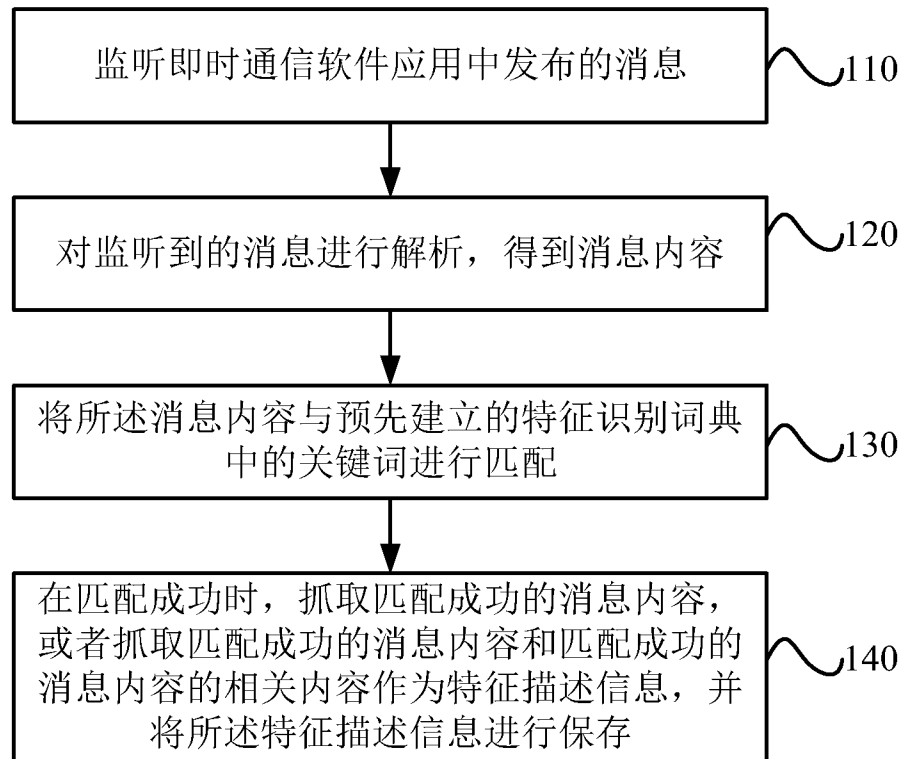


图 1

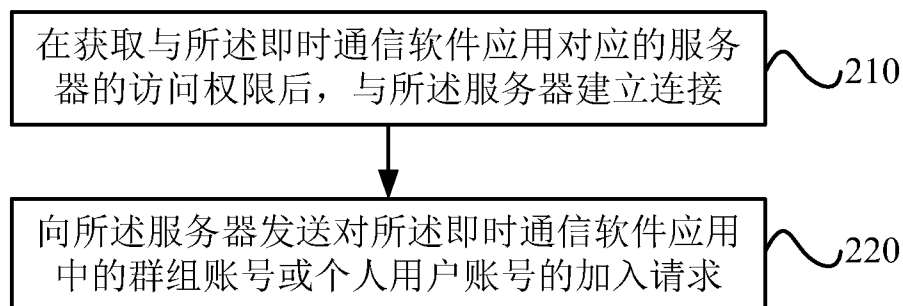


图 2

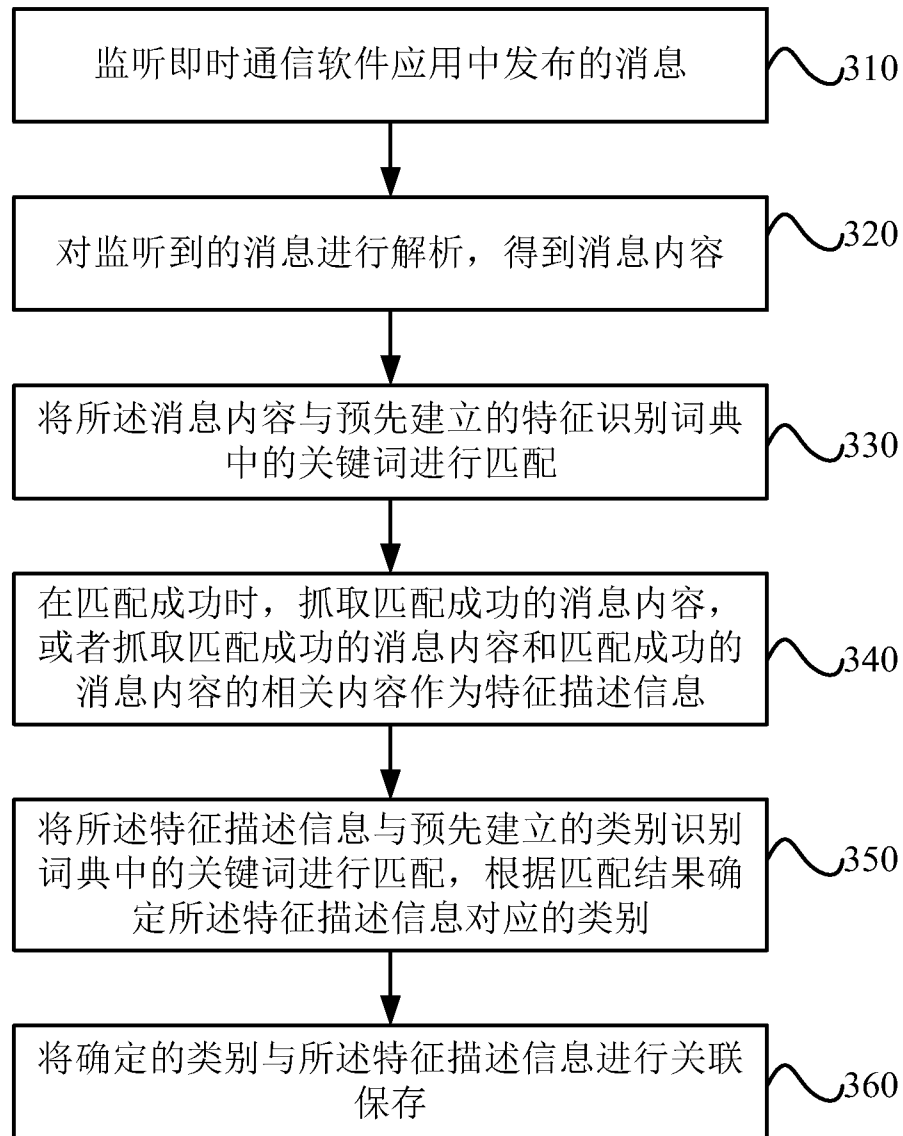


图 3a

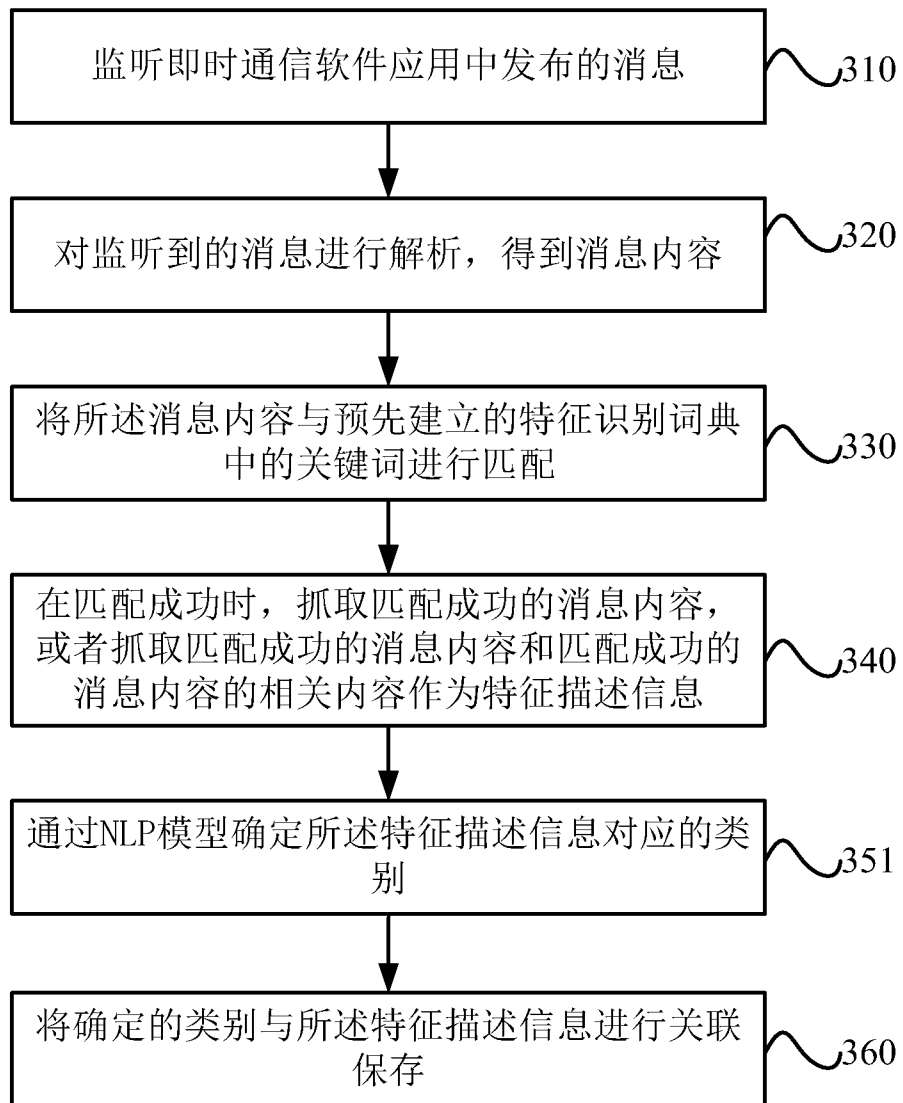


图 3b

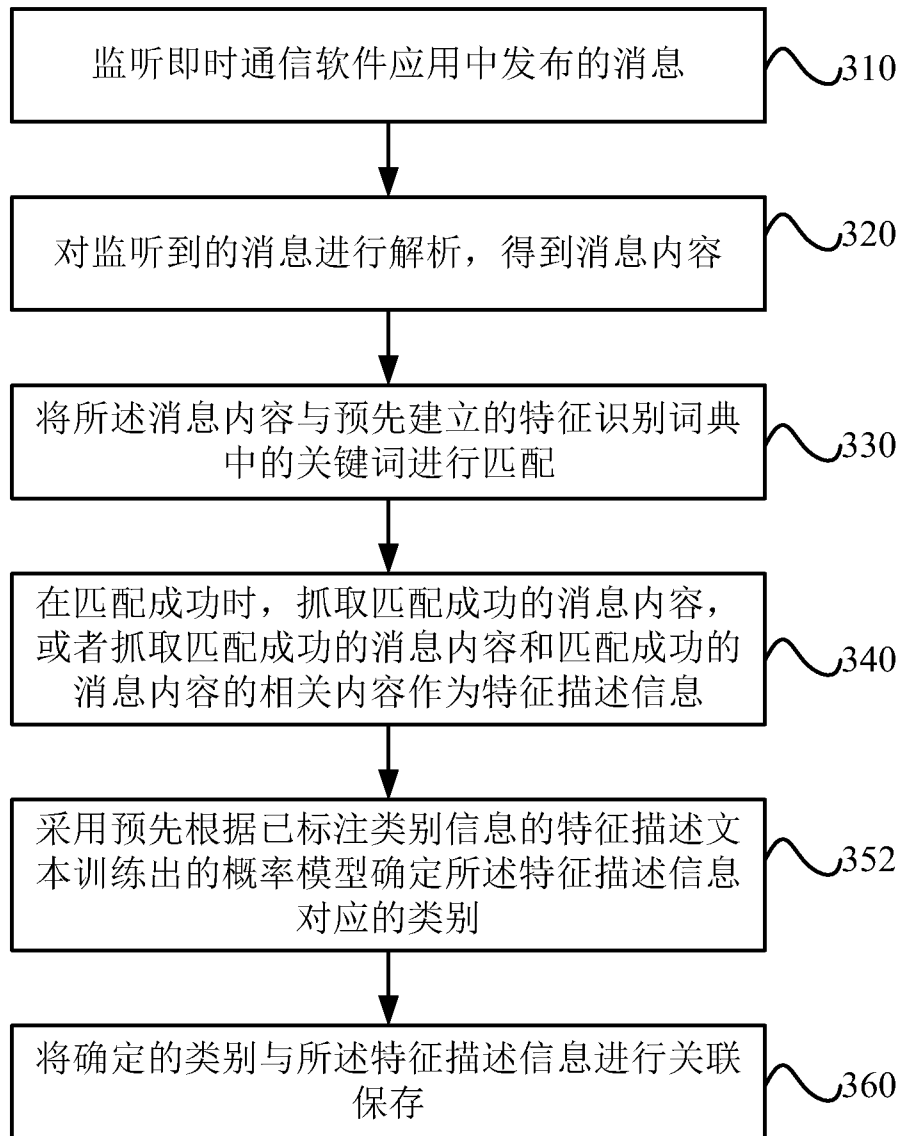


图 3c

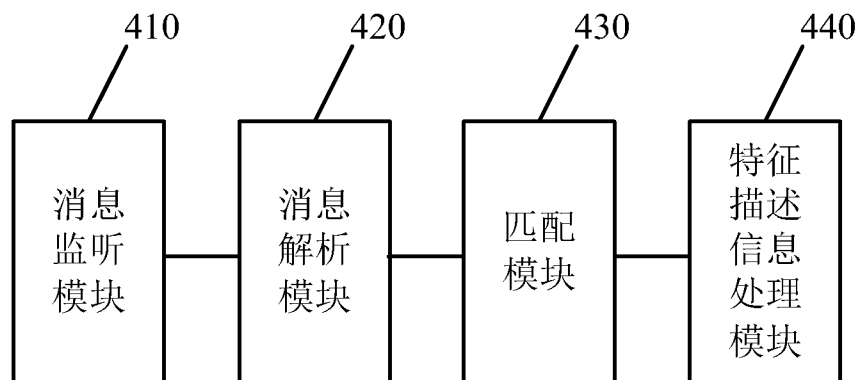


图 4

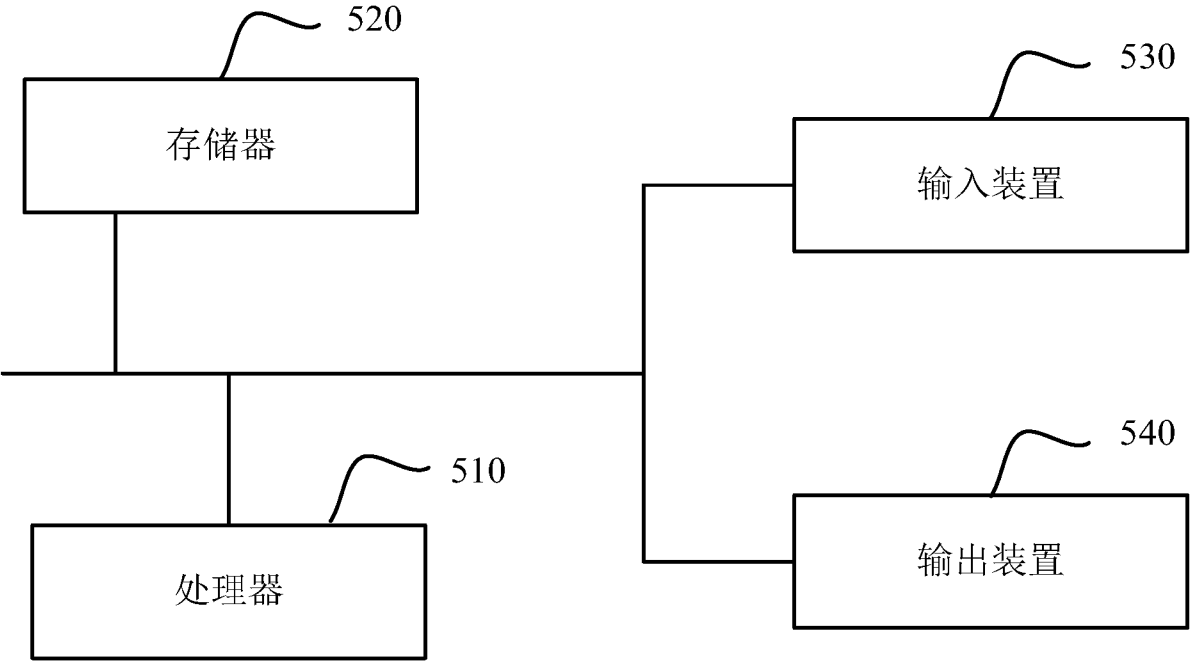


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/086095

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

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: instant communication, monitor+, message+, parse+, analys+, characteristic, identify, dictionary, key word+, match+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104346480 A (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD) 11 February 2015 (11.02.2015) claims 1 to 18	1-18
X	CN 103605690 A (BEIJING QIHOO TECHNOLOGY CO et al.) 26 February 2014 (26.02.2014) description, paragraphs [0017] to [0020] and figures 1 to 6	1-19
A	CN 1987852 A (TENGXUN SCIENCE & TECHNOLOGY) 27 June 2007 (27.06.2007) the whole document	1-19
A	CN 102323933 A (ZHANG, Xiao) 18 January 2012 (18.01.2012) the whole document	1-19
A	US 2002133477 A1 (ABEL, GLENN) 19 September 2002 (19.09.2002) the whole document	1-19

☐ 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 08 October 2015	Date of mailing of the international search report 29 October 2015
Name and mailing address of the ISA 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 BAI, Lushuang Telephone No. (86-10) 82246937

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2015/086095

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104346480 A	11 February 2015	None	
CN 103605690 A	26 February 2014	WO 2015062377 A1	07 May 2015
CN 1987852 A	27 June 2007	None	
CN 102323933 A	18 January 2012	None	
US 2002133477 A1	19 September 2002	None	

国际检索报告

国际申请号

PCT/CN2015/086095

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

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, EPODOC: 即时通信, 监听, 消息, 解析, 分析, 特征, 识别, 词典, 关键词, 匹配, instant communication, monitor+, message+, parse+, analys+, characteristic, dictionary, key word+, match+

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
PX	CN 104346480 A (百度在线网络技术北京有限公司) 2015年 2月 11日 (2015 - 02 - 11) 权利要求1-18	1-18
X	CN 103605690 A (北京奇虎科技有限公司 等) 2014年 2月 26日 (2014 - 02 - 26) 说明书第[0017]-[0020]段, 附图1-6	1-19
A	CN 1987852 A (腾讯科技深圳有限公司) 2007年 6月 27日 (2007 - 06 - 27) 全文	1-19
A	CN 102323933 A (张潇) 2012年 1月 18日 (2012 - 01 - 18) 全文	1-19
A	US 2002133477 A1 (ABEL, GLENN) 2002年 9月 19日 (2002 - 09 - 19) 全文	1-19

☐ 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

2015年 10月 8日

国际检索报告邮寄日期

2015年 10月 29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)
北京市海淀区蓟门桥西土城路6号
100088 中国

传真号 (86-10)62019451

授权官员

白露霜

电话号码 (86-10)82246937

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/086095

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	104346480	A	2015年 2月 11日	无			
CN	103605690	A	2014年 2月 26日	WO	2015062377	A1	2015年 5月 7日
CN	1987852	A	2007年 6月 27日	无			
CN	102323933	A	2012年 1月 18日	无			
US	2002133477	A1	2002年 9月 19日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)