

(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 3 月 9 日 (09.03.2017)



(10) 国际公布号
WO 2017/036047 A1

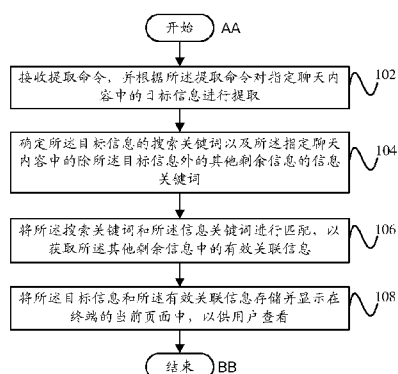
- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/100219
- (22) 国际申请日: 2015 年 12 月 31 日 (31.12.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510549296.7 2015 年 8 月 31 日 (31.08.2015) CN
- (71) 申请人: 宇龙计算机通信科技(深圳)有限公司
(YULONG COMPUTER TELECOMMUNICATION
SCIENTIFIC (SHENZHEN) CO., LTD.) [CN/CN]; 中国
广东省深圳市南山区科技园北区梦溪道 2 号,
Guangdong 518057 (CN)。
- (72) 发明人: 徐宇杰 (XU, Yujie); 中国广东省深圳市南
山区科技园北区梦溪道 2 号, Guangdong 518057
(CN)。 杨中云 (YANG, Zhongyun); 中国广东省深
圳市南山区科技园北区梦溪道 2 号, Guangdong
518057 (CN)。 沈小刚 (SHEN, Xiaogang); 中国广东
省深圳市南山区科技园北区梦溪道 2 号, Guangdong
518057 (CN)。

- (74) 代理人: 广州三环专利代理有限公司 (GUANG-
ZHOU SCIHEAD PATENT AGENT CO., LTD.); 中
国广东省广州市越秀区先烈中路 80 号汇华商贸大
厦 1508 室, Guangdong 510070 (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 EXTRACTION METHOD AND INFORMATION EXTRACTION DEVICE

(54) 发明名称: 信息提取方法和信息提取装置



- 101 Receive an extraction instruction, and extract, according to the extraction instruction, target information from specific chat content
- 104 Determine a search keyword of the target information and an information keyword of the other remaining information excluding the target information in the specific chat content
- 106 Perform matching based on the search keyword and the information keyword to obtain valid related information in the other remaining information
- 108 Store the target information and the valid related information, and display the same in a current page of a terminal for a user to view
- AA Start
- BB End

图 1

(57) Abstract: An information extraction method and information extraction device. The information extraction method comprises: receiving an extraction instruction, and extracting, according to the extraction instruction, target information from specific chat content (102); determining a search keyword of the target information and an information keyword of the other remaining information excluding the target information in the specific chat content (104); performing matching based on the search keyword and the information keyword to obtain valid related information in the other remaining information (106); and storing the target information and the valid related information, and displaying the same in a current page of a terminal for a user to view (108). The present invention effectively improves, by using the method and device, efficiency of searching chat information, such that a user can conveniently and quickly find valid chat information in a huge amount of redundant chat information, and also prevents wasting a great deal of data usage quota on browsing chat history records, thus improving user experience.

(57) 摘要:

[见续页]



WO 2017/036047 A1

**根据细则 4.17 的声明:**

— 关于申请人有权要求在先申请的优先权(细则 4.17(iii))

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种信息提取方法、一种信息提取装置，其中，所述信息提取方法包括：接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取（102）；确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息的信息关键词（104）；将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息（106）；将所述目标信息和所述有效关联信息存储并显示在终端的当前页面中，以供用户查看（108）。通过该方法和装置，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

信息提取方法和信息提取装置

本申请要求于 2015 年 08 月 31 日提交中国专利局，申请号为 201510549296.7、发明名称为“信息提取方法和信息提取装置”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。

5

技术领域

本发明涉及终端技术领域，具体而言，涉及一种信息提取方法和一种信息提取装置。

10 背景技术

目前，随着日常生活中智能移动终端（如手机）的大量普及，利用智能移动终端进行聊天也越来越便利，但是，利用智能移动终端聊天过程中会产生大量冗余、无效的信息，尤其在多人聊天模式下，问题与回答中间会夹杂其他无用的信息，从而无法分辨与某一个人的聊天信息。

15 而相关技术中，面对大量冗余、无效的信息占据大量屏幕空间而需要获取有效信息时，一般都需要用户翻阅聊天记录，逐条查阅，甚至，信息量过大时，需要用户连接网络获取服务器上的缓存信息以进行历史信息检索。如此，不仅降低了信息检索效率，还会大大增加用户数据流量的消耗，给用户带来了不好的体验。

20 因此，如何提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验成为亟待解决的问题。

发明内容

25 本发明正是基于上述问题，提出了一种新的技术方案，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

有鉴于此，本发明的第一方面提出了一种信息提取方法，包括：接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息的信息关键词；将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；将所述目标信息和所述有效关联信息存储并显示在终端的当前页面中，以供用户查看。

在该技术方案中，通过根据提取命令对指定聊天内容中的目标信息进行提取，以供确定目标信息的搜索关键词，同时确定筛选出的该指定聊天内容中除该目标信息以外的其他剩余信息的信息关键词，然后将提取到的目标信息的搜索关键词和其他剩余信息的信息关键词进行匹配，用以获取其他剩余信息中与搜索关键词相匹配的有效关联信息，并将目标信息和有效关联信息存储并显示在终端的当前页面中，优选地，可以设置显示在指定的下拉标签中，在需要查看时点击下拉标签进行查看，操作便捷且不占用额外的屏幕空间，当然也可以设置其他显示方式进行显示，比如将目标信息与用户问题信息、有效答复信息分开显示等，如此，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

在上述技术方案中，优选地，所述接收提取命令，对指定聊天内容中的目标信息进行提取，具体包括：根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

在该技术方案中，可以根据用户选择命令提取确定指定聊天内容中的目标信息，即可以根据用户指令收集重要的信息，例如，用户可以通过手动记录的方式，将用户主观认为重要的信息进行收集，和/或根据预设提取模型对指定聊天内容中的目标信息进行提取，其中，预设提取模型可以是指带有问号、带有疑问副词（比如，什么、怎么、吗）和/或用户自主设置的问题模型，当检测到聊天内容中含有与预设问题模型匹配的句子时，即可判定其为目标信息，从而完成对目标信息准确有效地提取，而且通过用户主观选择和根据预设提取模型进行提取目标信息，可以有效地互补，对

提取的有效信息进行查缺补漏，便于用户查看，进而使用户具有最佳的使用体验。

在上述技术方案中，优选地，按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供
5 所述用户查看。

在该技术方案中，将存储在当前页面中的目标信息和有效关联信息优选地可以按照该指定聊天内容生成的时间顺序进行显示，以便将最近的目标信息和有效关联信息显示在最下面，从而使用户查看时更加清晰、明了。

在上述技术方案中，优选地，根据预定提取规则确定所述目标信息的
10 所述搜索关键词以及所述其他剩余信息中的所述信息关键词。

在该技术方案中，根据预定提取规则确定目标信息的搜索关键词以及其他剩余信息中的信息关键词，例如，预定提取规则可以为将目标信息中的名词和/或形容词作为关键词，根据句子中词语的词性作为依据确定关键词，当然，预定提取规则还可以是提取目标信息中有关某一领域的词语作为
15 关键词，例如，涉及地理山川领域的地理名称、计算机领域的硬件信息或者算法、日常生活领域的食物名称、交通工具等，根据不同领域的相关的词语作为依据确定关键词。

在上述技术方案中，优选地，所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；所
20 述目标信息包括：用户问题信息。

在该技术方案中，指定聊天内容包括但不限于当前多人聊天内容中的文字信息和语音信息、最新的缓存未读聊天内容的文字信息和语音信息和/或历史聊天记录中的文字信息和语音信息，通过对文字信息和语音信息中的目标信息和有效关联信息的提取，从而可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，
25 并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

根据本发明的第二个方面提出了一种信息提取装置，用于终端，包括：提取模块，用于接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；确定模块，用于确定所述目标信息的搜索关键词以及

所述指定聊天内容中的除所述目标信息外的其他剩余信息其他剩余信息的信息关键词；匹配模块，用于将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；显示模块，用于将所述目标信息和所述有效关联信息存储并显示在所述终端的当前页面中，以供用户查看。

在该技术方案中，通过根据提取命令对指定聊天内容中的目标信息进行提取，以供确定目标信息的搜索关键词，同时确定筛选出的该指定聊天内容中除该目标信息以外的其他剩余信息的信息关键词，然后将提取到的目标信息的搜索关键词和其他剩余信息的信息关键词进行匹配，用以获取其他剩余信息中与搜索关键词相匹配的有效关联信息，并将目标信息和有效关联信息存储并显示在终端的当前页面中，优选地，可以设置显示在指定的下拉标签中，在需要查看时点击下拉标签进行查看，操作便捷且不占用额外的屏幕空间，当然也可以设置其他显示方式进行显示，比如将目标信息与用户问题信息、有效答复信息分开显示等，如此，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

在上述技术方案中，优选地，所述提取模块具体用于：所述提取模块具体用于：根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

在该技术方案中，可以根据用户选择命令提取确定指定聊天内容中的目标信息，即可以根据用户指令收集重要的信息，例如，用户可以通过手动记录的方式，将用户主观认为重要的信息进行收集，和/或根据预设提取模型对指定聊天内容中的目标信息进行提取，其中，预设提取模型可以是指带有问号、带有疑问副词（比如，什么、怎么、吗）和/或用户自主设置的问题模型，当检测到聊天内容中含有与预设问题模型匹配的句子时，即可判定其为目标信息，从而完成对目标信息准确有效地提取，而且通过用户主观选择和根据预设提取模型进行提取目标信息，可以有效地互补，对提取的有效信息进行查缺补漏，便于用户查看，进而使用户具有最佳的使

用体验。

在上述技术方案中，优选地，所述显示模块具体用于：按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供所述用户查看。

5 在该技术方案中，将存储在当前页面中的目标信息和有效关联信息优选地可以按照该指定聊天内容生成的时间顺序进行显示，以便将最近的目标信息和有效关联信息显示在最下面，从而使用户查看时更加清晰、明了。

在上述技术方案中，优选地，所述确定模块具体用于：根据预定提取规则确定所述目标信息的所述搜索关键词以及所述其他剩余信息中的所述
10 信息关键词。

在该技术方案中，根据预定提取规则确定目标信息的搜索关键词以及其他剩余信息中的信息关键词，例如，预定提取规则可以为将目标信息中的名词和/或形容词作为关键词，根据句子中词语的词性作为依据确定关键词，当然，预定提取规则还可以是提取目标信息中有关某一领域的词语作为
15 关键词，例如，涉及地理山川领域的地理名称、计算机领域的硬件信息或者算法、日常生活领域的食物名称、交通工具等，根据不同领域的相关的词语作为依据确定关键词。

在上述技术方案中，优选地，所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；所
20 述目标信息包括：用户问题信息

在该技术方案中，指定聊天内容包括但不限于当前多人聊天内容中的文字信息和语音信息、最新的缓存未读聊天内容的文字信息和语音信息和/或历史聊天记录中的文字信息和语音信息，通过对文字信息和语音信息中的目标信息和有效关联信息的提取，从而可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，
25 并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

本发明的第三个方面提出了一种终端，包括上述技术方案中任一项所述的信息提取装置，因此，该终端具有和上述技术方案中任一项所述的信息提取装置相同的技术效果，在此不再赘述。

通过本发明的技术方案，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

5 附图说明

图 1 示出了根据本发明的一个实施例的信息提取方法的流程示意图；

图 2 示出了根据本发明的一个实施例的信息提取装置的结构示意图；

图 3 示出了根据本发明的一个实施例的终端的结构示意图；

图 4 示出了根据本发明的另一个实施例的信息提取方法的流程示意图；

图 5 示出了根据本发明的一个实施例的提取用户聊天信息的示意图。

具体实施方式

为了可以更清楚地理解本发明的上述目的、特征和优点，下面结合附图和具体实施方式对本发明进行进一步的详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本发明，但是，本发明还可以采用其他不同于在此描述的方式来实施，因此，本发明的保护范围并不受下面公开的具体实施例的限制。

图 1 示出了根据本发明的一个实施例的信息提取方法的流程示意图。

如图 1 所示，根据本发明的一个实施例的信息提取方法，包括：步骤 102，接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；步骤 104，确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息的信息关键词；步骤 106，将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；步骤 108，将所述目标信息和所述有效关联信息存储并显示在终端的当前页面中，以供用户查看。

在该技术方案中，通过根据提取命令对指定聊天内容中的目标信息进行提取，以供确定目标信息的搜索关键词，同时确定筛选出的该指定聊天

内容中除该目标信息以外的其他剩余信息的信息关键词，然后将提取到的目标信息的搜索关键词和其他剩余信息的信息关键词进行匹配，用以获取其他剩余信息中与搜索关键词相匹配的有效关联信息，并将目标信息和有效关联信息存储并显示在终端的当前页面中，优选地，可以设置显示在指定的下拉标签中，在需要查看时点击下拉标签进行查看，操作便捷且不占用额外的屏幕空间，当然也可以设置其他显示方式进行显示，比如将目标信息与用户问题信息、有效答复信息分开显示等，如此，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

在上述技术方案中，优选地，所述步骤 102 具体包括：根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

在该技术方案中，可以根据用户选择命令提取确定指定聊天内容中的目标信息，即可以根据用户指令收集重要的信息，例如，用户可以通过手动记录的方式，将用户主观认为重要的信息进行收集，和/或根据预设提取模型对指定聊天内容中的目标信息进行提取，其中，预设提取模型可以是指带有问号、带有疑问副词（比如，什么、怎么、吗）和/或用户自主设置的问题模型，当检测到聊天内容中含有与预设问题模型匹配的句子时，即可判定其为目标信息，从而完成对目标信息准确有效地提取，而且通过用户主观选择和根据预设提取模型进行提取目标信息，可以有效地互补，对提取的有效信息进行查缺补漏，便于用户查看，进而使用户具有最佳的使用体验。

在上述技术方案中，优选地，按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供所述用户查看。

在该技术方案中，将存储在当前页面中的目标信息和有效关联信息优选地可以按照该指定聊天内容生成的时间顺序进行显示，以便将最近的目标信息和有效关联信息显示在最下面，从而使用户查看时更加清晰、明了。

在上述技术方案中，优选地，根据预定提取规则确定所述目标信息的所述搜索关键词以及所述其他剩余信息中的所述信息关键词。

在该技术方案中，根据预定提取规则确定目标信息的搜索关键词以及其他剩余信息中的信息关键词，例如，预定提取规则可以为将目标信息中的名词和/或形容词作为关键词，根据句子中词语的词性作为依据确定关键词，当然，预定提取规则还可以是提取目标信息中有关某一领域的词语作为关键词，例如，涉及地理山川领域的地理名称、计算机领域的硬件信息或者算法、日常生活领域的食物名称、交通工具等，根据不同领域的相关的词语作为依据确定关键词。

在上述技术方案中，优选地，所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；所述目标信息包括：用户问题信息。

在该技术方案中，指定聊天内容包括但不限于当前多人聊天内容中的文字信息和语音信息、最新的缓存未读聊天内容的文字信息和语音信息和/或历史聊天记录中的文字信息和语音信息，通过对文字信息和语音信息中的目标信息和有效关联信息的提取，从而可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

图 2 示出了根据本发明的一个实施例的信息提取装置的结构示意图。

如图 2 所示，根据本发明的一个实施例的信息提取装置 200，包括：提取模块 202，提取模块，用于接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；确定模块 204，用于确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息其他剩余信息的信息关键词；匹配模块 206，用于将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；显示模块 208，用于将所述目标信息和所述有效关联信息存储并显示在所述终端的当前页面中，以供用户查看。

在该技术方案中，通过根据提取命令对指定聊天内容中的目标信息进行提取，以供确定目标信息的搜索关键词，同时确定筛选出的该指定聊天

内容中除该目标信息以外的其他剩余信息的信息关键词，然后将提取到的目标信息的搜索关键词和其他剩余信息的信息关键词进行匹配，用以获取其他剩余信息中与搜索关键词相匹配的有效关联信息，并将目标信息和有效关联信息存储并显示在终端的当前页面中，优选地，可以设置显示在指定的下拉标签中，在需要查看时点击下拉标签进行查看，操作便捷且不占用额外的屏幕空间，当然也可以设置其他显示方式进行显示，比如将目标信息与用户问题信息、有效答复信息分开显示等，如此，可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

在上述技术方案中，优选地，所述提取模块 202 具体用于：所述提取模块具体用于：根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

在该技术方案中，可以根据用户选择命令提取确定指定聊天内容中的目标信息，即可以根据用户指令收集重要的信息，例如，用户可以通过手动记录的方式，将用户主观认为重要的信息进行收集，和/或根据预设提取模型对指定聊天内容中的目标信息进行提取，其中，预设提取模型可以是指带有问号、带有疑问副词（比如，什么、怎么、吗）和/或用户自主设置的问题模型，当检测到聊天内容中含有与预设问题模型匹配的句子时，即可判定其为目标信息，从而完成对目标信息准确有效地提取，而且通过用户主观选择和根据预设提取模型进行提取目标信息，可以有效地互补，对提取的有效信息进行查缺补漏，便于用户查看，进而使用户具有最佳的使用体验。

在上述技术方案中，优选地，所述显示模块 208 具体用于：按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供所述用户查看。

在该技术方案中，将存储在当前页面中的目标信息和有效关联信息优选地可以按照该指定聊天内容生成的时间顺序进行显示，以便将最近的目标信息和有效关联信息显示在最下面，从而使用户查看时更加清晰、明了。

在上述技术方案中，优选地，所述确定模块 204 具体用于：根据预定提取规则确定所述目标信息的所述搜索关键词以及所述其他剩余信息中的所述信息关键词。

5 在该技术方案中，根据预定提取规则确定目标信息的搜索关键词以及其他剩余信息中的信息关键词，例如，预定提取规则可以为将目标信息中的名词和/或形容词作为关键词，根据句子中词语的词性作为依据确定关键词，当然，预定提取规则还可以是提取目标信息中有关某一领域的词语作为关键词，例如，涉及地理山川领域的地理名称、计算机领域的硬件信息或者算法、日常生活领域的食物名称、交通工具等，根据不同领域的相关的
10 的词语作为依据确定关键词。

在上述技术方案中，优选地，所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；所述目标信息包括：用户问题信息

在该技术方案中，指定聊天内容包括但不限于当前多人聊天内容中的
15 文字信息和语音信息、最新的缓存未读聊天内容的文字信息和语音信息和/或历史聊天记录中的文字信息和语音信息，通过对文字信息和语音信息中的目标信息和有效关联信息的提取，从而可以有效地提高聊天信息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

20 图 3 示出了根据本发明的一个实施例的终端的结构示意图。

如图 3 所示，根据本发明的一个实施例的终端 300，包括上述技术方案中任一项所述的信息提取装置 200，因此，该终端 300 具有和上述技术方案中任一项所述的信息提取装置 200 相同的技术效果，在此不再赘述。

下面结合图 4 和图 5 具体说明本发明的技术方案，图 4 示出了根据本
25 发明的另一个实施例的信息提取方法的流程示意图；图 5 示出了根据本发明的一个实施例的智能收集并提取用户聊天信息的示意图；具体地以目标信息为用户问题信息为例进行说明。

根据本发明的一个实施例的智能收集并提取用户聊天信息，其实现原理为：首先，监测当前用户输入的文字、语音等信息，自动收集（获取）

问题类信息（目标信息），如“A 型号手机，性能咋样？”，当然，也可以通过手动记录的方式，将用户主观认为重要的信息进行收集（即根据用户的选择命令），此类信息未必就是相应用户问题信息的解决方案，甚至是不符合问题逻辑的，如“B 型号手机价比高啊”，然后根据用户的回答，

5 对用户答复的信息进行收集和提取重要的信息，最后生成一张“聊天信息精简”的写字板模式（下拉标签显示），在用户聊天结束后，可直接看聊天信息精简即可。具体地实现步骤，如图 4 所示：

第一步，进入聊天模式后，比如腾讯 QQ 或微信等的多人聊天模式，开启智能提取信息功能，对用户的问题及用户认为重要的信息进行收集和

10 提取；

第二步，问题关键字提取模型，对用户提问进行关键字（搜索关键词）提取，如图 5，提取的关键字为“A 型号手机”；

第三步，信息关键字提取模型，检测聊天信息中的回答信息，进行信息收集，并提取有效信息（其他剩余信息）；

15 第四步，将问题关键字和答复关键字（信息关键词）进行匹配，提取出有效的答复信息（有效关联信息），如图 5，“A 型号新出的，曲屏，不错”，其中关键字“A 型号”就和上文的问题关键字“A 型号手机”匹配上了；

第五步，关键信息收集模型，对已经匹配上的问题与有效答复，进行

20 收集，将用户的答复提取并放入信息收集板（下拉标签）中，如图 5，右侧的信息提取和收集模板；

第六步，用户聊天结束后，可自行关闭信息收集提取模式，或者在聊天过程中，手动关闭即可；

第七步，退出聊天信息收集和提取模式。

25 在该技术方案中，可以去掉聊天信息，尤其是群聊模式中的冗余信息或未能及时查看的新生成的聊天信息，收集并提取有用的聊天信息，同时避免用户浏览历史聊天信息时，浪费数据流量在上下翻阅缓存聊天记录上，以提高聊天信息检索的高效型、可操作性以及便捷性。

以上结合附图详细说明了本发明的技术方案，可以有效地提高聊天信

息检索的效率，使用户在大量冗余聊天信息中可以方便、快速地找到有效聊天信息，并避免消耗大量的数据流量浏览历史聊天记录，从而提升用户体验。

5 以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求

1. 一种信息提取方法，其特征在于，包括：

5 接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；

确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息的信息关键词；

将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；

10 将所述目标信息和所述有效关联信息存储并显示在终端的当前页面中，以供用户查看。

2. 根据权利要求 1 所述的信息提取方法，其特征在于，所述接收提取命令，对指定聊天内容中的目标信息进行提取，具体包括：

15 根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

3. 根据权利要求 2 所述的信息提取方法，其特征在于，

按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供所述用户查看。

20 4. 根据权利要求 1 至 3 中任一项所述的信息提取方法，其特征在于，根据预定提取规则确定所述目标信息的所述搜索关键词以及所述其他剩余信息中的所述信息关键词。

5. 根据权利要求 1 至 3 中任一项所述的信息提取方法，其特征在于，

所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；

25 所述目标信息包括：用户问题信息。

6. 一种信息提取装置，其特征在于，包括：

提取模块，用于接收提取命令，并根据所述提取命令对指定聊天内容中的目标信息进行提取；

确定模块，用于确定所述目标信息的搜索关键词以及所述指定聊天内容中的除所述目标信息外的其他剩余信息其他剩余信息的信息关键词；

匹配模块，用于将所述搜索关键词和所述信息关键词进行匹配，以获取所述其他剩余信息中的有效关联信息；

5 显示模块，用于将所述目标信息和所述有效关联信息存储并显示在所述终端的当前页面中，以供用户查看。

7. 根据权利要求 6 所述的信息提取装置，其特征在于，所述提取模块具体用于：

10 根据接收到的用户选择命令和/或预设提取模型提取所述指定聊天内容中的所述目标信息。

8. 根据权利要求 7 所述的信息提取装置，其特征在于，所述显示模块具体用于：

按照所述指定聊天内容生成的时间顺序将所述目标信息和所述有效关联信息存储并显示在所述当前页面中，以供所述用户查看。

15 9. 根据权利要求 6 至 8 中任一项所述的信息提取装置，其特征在于，所述确定模块具体用于：

根据预定提取规则确定所述目标信息的所述搜索关键词以及所述其他剩余信息中的所述信息关键词。

20 10. 根据权利要求 6 至 8 中任一项所述的信息提取装置，其特征在于，所述指定聊天内容包括：当前多人聊天内容、缓存未读聊天内容和/或历史聊天记录中的文字信息和语音信息；

所述目标信息包括：用户问题信息。

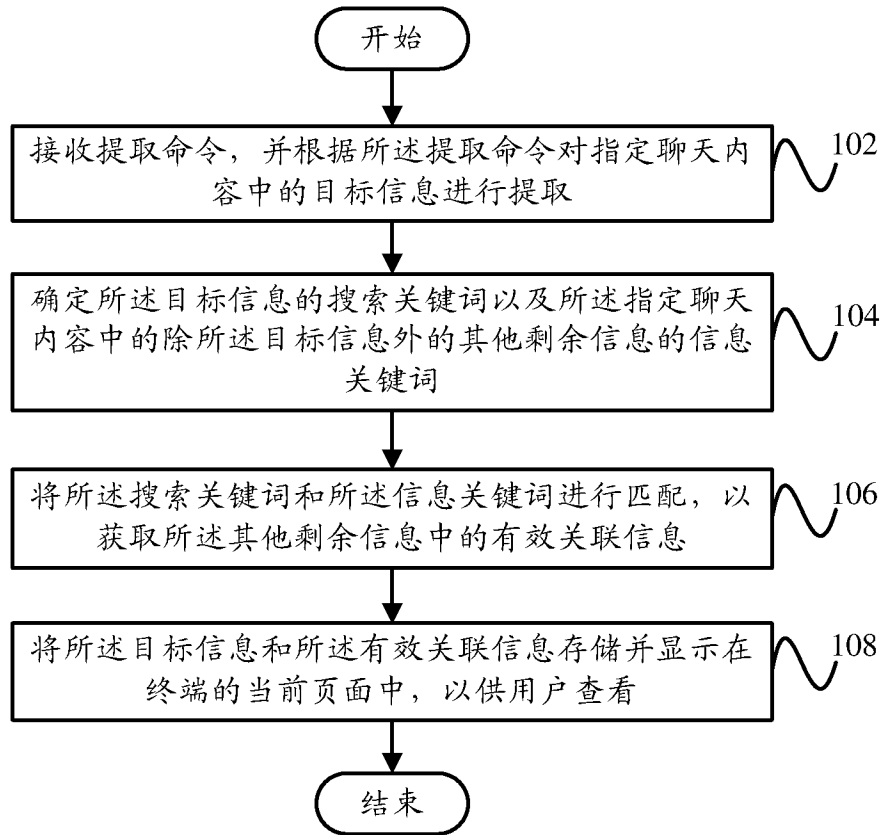


图 1

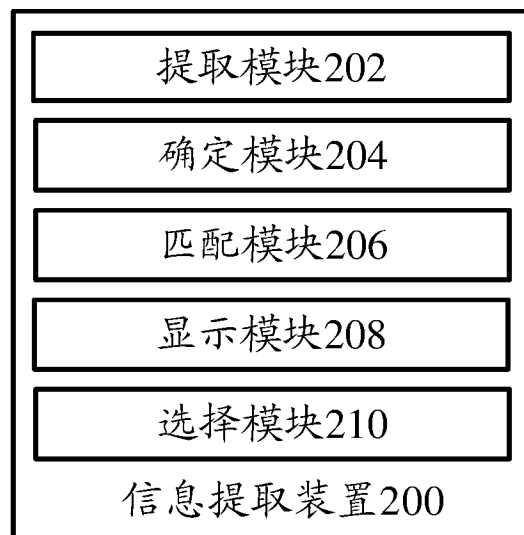


图 2



图 3

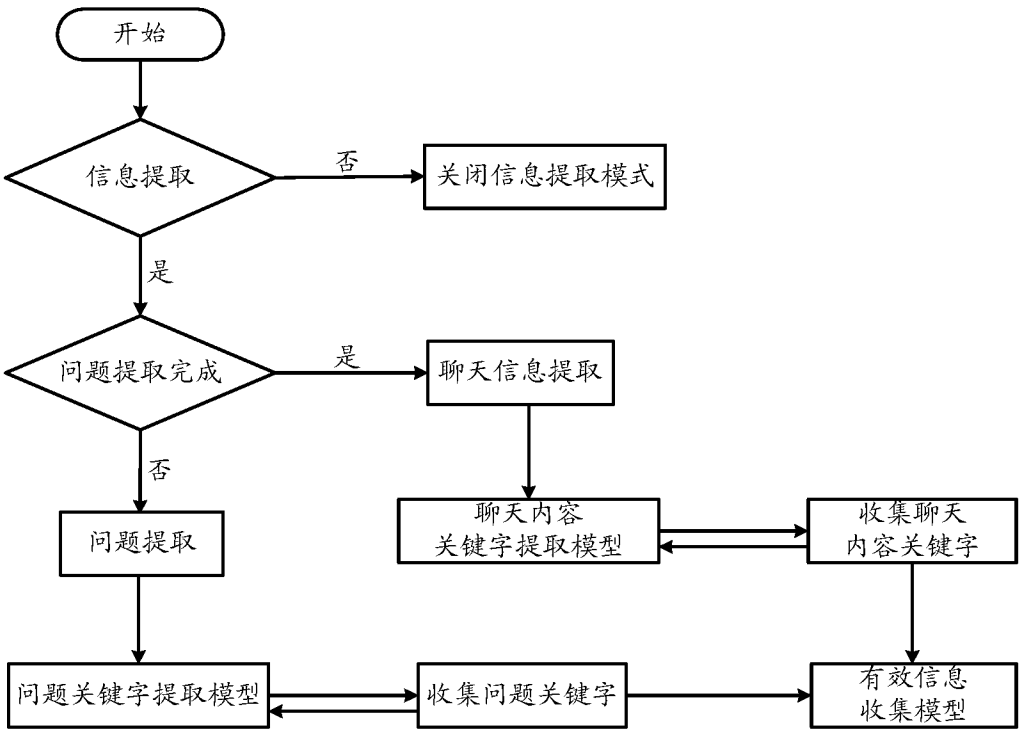


图 4

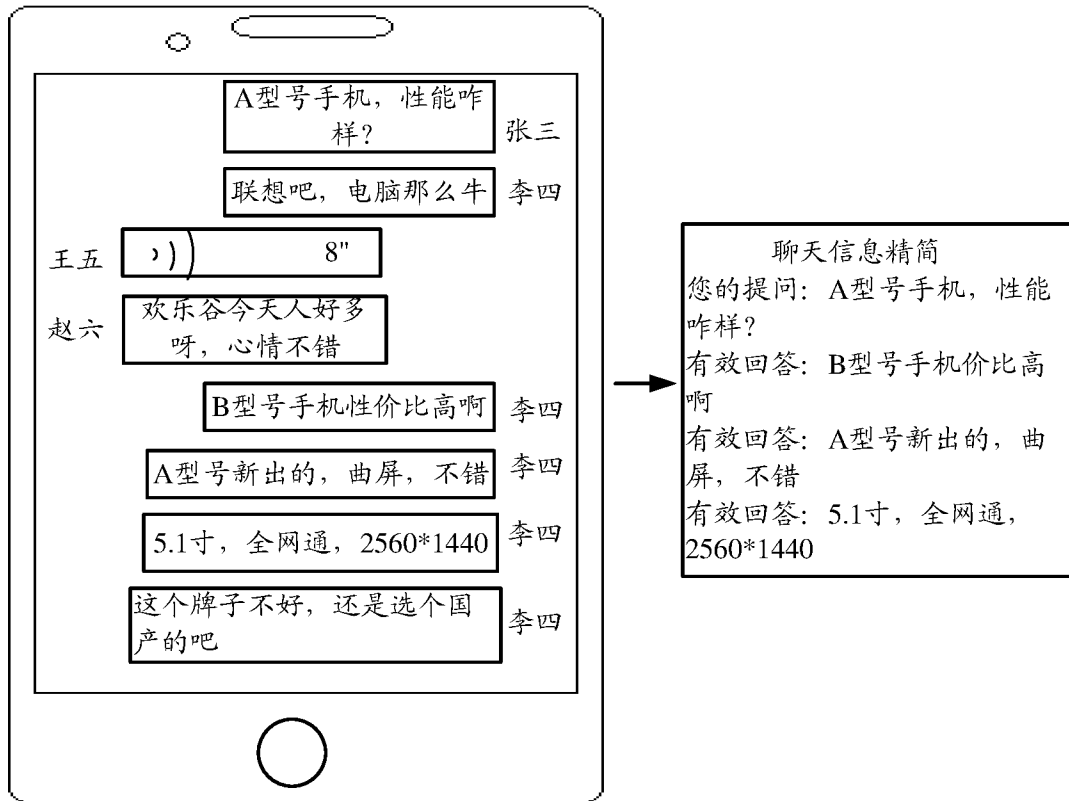


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/100219

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)

WPI, EPODOC, CNPAT, CNKI, GOOGLE, IEEE: chat, talk, keyword, extract, match, relation, program, search, record, history

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104361003 A (GOLDEN VAST MACAO COMMERCIAL OFFSHORE LIMITED), 18 February 2015 (18.02.2015), description, paragraphs [0050]-[0055] and [0078]-[0080]	1-10
A	CN 101072205 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 14 November 2007 (14.11.2007), the whole document	1-10
A	CN 104133908 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.), 05 November 2014 (05.11.2014), the whole document	1-10
A	US 2014067842 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 06 March 2014 (06.03.2014), the whole document	1-10
A	CN 104142964 A (TENCENT TECHNOLOGY (BEIJING) CO., LTD.), 12 November 2014 (12.11.2014), the whole document	1-10

☐ 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 24 March 2016 (24.03.2016)	Date of mailing of the international search report 27 May 2016 (27.05.2016)
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 ZHANG, Tao Telephone No.: (86-10) 62414427

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/100219

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104361003 A	18 February 2015	None	
CN 101072205 A	14 November 2007	WO 2008154832 A1	24 December 2008
		VN 22617 A	26 April 2010
		IN CHENP201000255 E	16 July 2010
CN 104133908 A	05 November 2014	None	
US 2014067842 A1	06 March 2014	CN 103678269 A	26 March 2014
CN 104142964 A	12 November 2014	WO 2014198132 A1	18 December 2014
		US 2014372403 A1	18 December 2014

国际检索报告

国际申请号

PCT/CN2015/100219

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))

WPI, EPODOC, CNPAT, CNKI, GOOGLE, IEEE 聊天, 对话, 谈话, 关键词, 提取, 匹配, 关联, 程序, 搜索, 记录, 历史, chat, talk, keyword, extract, match, relation, program, search, record, history

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 104361003 A (金硕澳门离岸商业服务有限公司) 2015年 2月 18日 (2015 - 02 - 18) 说明书第[0050]-[0055]段, 第[0078]-[0080]段	1-10
A	CN 101072205 A (腾讯科技深圳有限公司) 2007年 11月 14日 (2007 - 11 - 14) 全文	1-10
A	CN 104133908 A (北京奇虎科技有限公司 等) 2014年 11月 5日 (2014 - 11 - 05) 全文	1-10
A	US 2014067842 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 3月 6日 (2014 - 03 - 06) 全文	1-10
A	CN 104142964 A (腾讯科技北京有限公司) 2014年 11月 12日 (2014 - 11 - 12) 全文	1-10

☐ 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

2016年 3月 24日

国际检索报告邮寄日期

2016年 5月 27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

授权官员

张涛

传真号 (86-10)62019451

电话号码 (86-10)62414427

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/100219

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	104361003	A	2015年 2月 18日	无			
CN	101072205	A	2007年 11月 14日	WO	2008154832	A1	2008年 12月 24日
				VN	22617	A	2010年 4月 26日
				IN	CHENP201000255	E	2010年 7月 16日
CN	104133908	A	2014年 11月 5日	无			
US	2014067842	A1	2014年 3月 6日	CN	103678269	A	2014年 3月 26日
CN	104142964	A	2014年 11月 12日	WO	2014198132	A1	2014年 12月 18日
				US	2014372403	A1	2014年 12月 18日

表 PCT/ISA/210 (同族专利附件) (2009年7月)