

(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 4 月 18 日 (18.04.2019)



(10) 国际公布号
WO 2019/072007 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

路 969 号 3 号楼 5 楼 阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(21) 国际申请号: PCT/CN2018/099542

(22) 国际申请日: 2018 年 8 月 9 日 (09.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710946879.2 2017 年 10 月 12 日 (12.10.2017) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (**ALIBABA GROUP HOLDING LIMITED**) [—/CN]; 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 US): 王彦锋 (**WANG, Yanfeng**) [CN/CN]; 中国浙江省杭州市余杭区文一西

(74) 代理人: 北京博思佳知识产权代理有限公司 (**BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (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, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** DATA PROCESSING METHOD AND DEVICE

(54) 发明名称: 一种数据处理方法及装置

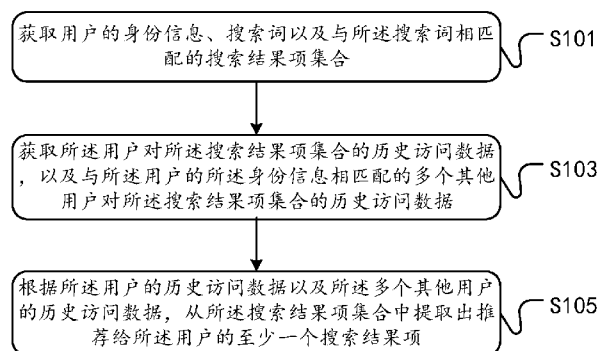


图 1

S101 Acquire identity information of a user, a search term, and a search result item set matching the search term
S103 Acquire historical access data of the user for the search result item set, and historical access data of a plurality of other users, matching the identity information of the user, for the search result item set
S105 According to the historical access data of the user and the historical access data of the plurality of other users, extract from the search result item set at least one search result item to be recommended to the user

(57) **Abstract:** A data processing method and device. The method comprises: acquiring identity information of a user, a search term, and a search result item set matching the search term (S101); acquiring historical access data of the user for the search result item set, and historical access data of a plurality of other users, matching the identity information of the user, for the search result item set (S103); and according to the historical access data of the user and the historical access data of the plurality of other users, extracting from the search result item set at least one search result item to be recommended to the user (S105). This solution enables the search result item recommended to the user to better meet the requirements of the user and the identity information of the user.

(57) **摘要:** 一种数据处理方法及装置。所述方法包括: 获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合 (S101); 获取所述用户对所述搜索结果项集合的历史访问数据, 以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据 (S103); 根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项 (S105)。利用本方案, 可以使得推荐给用户的搜索结果项更加符合用户的需求以及用户的身份信息。

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))。

一种数据处理方法及装置

技术领域

[01]本申请涉及数据处理技术领域，特别涉及一种数据处理方法及装置。

背景技术

5 [02]近几年，基于互联网平台提供的服务越来越多，其中，数据查询服务已经在各个行业中广泛应用。数据查询服务可以根据用户输入的搜索词从平台上获取相关数据，并将相关数据展示给用户，具体的相关数据例如可以包括文档、产品信息等数据。

[03]在一些特定行业内，如互联网行业，随着企业不断的发展，企业内信息系统沉淀积累的数据信息较多。由于企业内部数据信息的保密性和不可公开性，诸多信息一般不允
10 许对外部搜索引擎开放，仅允许在组织架构的局域网内访问。企业内部积累的数据信息对于企业用户来说，是具有价值的资料，对推动企业内部知识转移和学习具有重要意义。但是，目前很多行业内组织架构复杂，组织架构内用户在根据搜索词搜索企业内部数据时，企业内部组织架构往往仅根据搜索词匹配到与所述搜索词相似的内部数据，并将该数据展示给用户。

15 [04]利用上述方式获取的用户数据往往只能获取一些字面上与搜索词相关的数据信息，但是却无法满足用户的实际搜索诉求。因此，现有技术中亟需一种与用户实际的搜索需求相匹配的数据处理方式。

发明内容

[05]本申请实施例的目的在于提供一种数据处理方法及装置，可以使得推荐给用户的搜索
20 结果项更加符合用户的需求以及用户的身份信息。

[06]本申请实施例提供的一种数据处理方法及装置具体是这样实现的：

[07]一种数据处理方法，所述方法包括：获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；获取所述用户对所述搜索结果项集合的历史访问数据，以及
25 与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[08]一种数据处理装置，所述装置包括：结果项获取单元，用于获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；历史数据获取单元，用于获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；推荐结果项提取单元，用于根据
5 所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[09]一种数据处理装置，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器执行所述指令时实现：获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户
10 的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[10]本申请提供的数据处理方法及装置，可以根据用户的搜索词匹配到搜索结果项集合，再根据用户对所述搜索结果项集合的历史访问数据，以及与所述用户身份信息相匹配的
15 多个其他用户对所述搜索结果项集合的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。利用用户以及与用户身份信息相匹配的多个用户的历史数据，对推荐给用户的搜索结果项产生影响，使得推荐给用户的搜索结果项更加符合用户的需求以及用户的身份信息。尤其对于处于组织架构内的用户来说，用户在组织架构内搜索到的相关数据更加符合用户的搜索需求以及用户在组织架构内的身份。

20 附图说明

[11]为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

25 [12]图 1 是本申请提供的数据处理方法的一种实施例的流程示意图；

[13]图 2 是本申请提供的提取搜索结果项方法的一种实施例的方法流程示意图；

[14]图 3 是本申请提供的获取频繁项方法的一种实施例的方法流程图；

[15]图 4 是本申请提供的获取频繁项方法的另一种实施例的方法流程图；

[16]图 5 是本申请提供的推荐搜索结果项方法的一种实施例的方法流程图；

[17]图 6 是本申请提供的数据处理装置的一种实施例的模块结构示意图；

[18]图 7 是本申请提供的数据处理装置的另一种实施例的模块结构示意图。

具体实施方式

- 5 [19]为了使本技术领域的人员更好地理解本申请中的技术方案，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，所描述的实施例仅仅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作出创造性劳动前提下所获得的所有其他实施例，都应当属于本申请保护的范围。
- 10 [20]下面结合附图对本申请所述的数据处理方法进行详细的说明。图 1 是本申请提供的数据处理方法的一种实施例的方法流程示意图。虽然本申请提供了如下述实施例或附图所示的方法操作步骤，但基于常规或者无需创造性的劳动在所述方法中可以包括更多或者更少的操作步骤。在逻辑性上不存在必要因果关系的步骤中，这些步骤的执行顺序不限于本申请实施例提供的执行顺序。所述方法在实际中的数据处理过程中，可以按照实施例或者附图所示的方法顺序执行或者并行执行（例如并行处理器或者多线程处理的环境）。
- 15 [21]具体的本申请提供的一种数据处理方法的一种实施例如图 1 所示，所述方法可以包括：
- [22]S101：获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合。
- 20 [23]本说明书中，所述用户的身份信息可以包括所述用户在组织架构中的身份信息。所述组织架构例如可以包括企业、学校、图书馆等具有内部局域网的机构，在所述组织结构中，用户可以通过搜索词搜索组织架构提供的数据。例如，对于企业来说，用户的身份信息可以包括该用户所在的部门、职位（如管理人员、研发人员、文员等等）。对于学校来说，用户的身份信息可以包括该用户所在的学院、职位（教授、讲师、辅导员、
- 25 博士研究生、硕士研究生、本科生等等）。对于图书馆，可以具有与学校相似的身份信息划分，对此不再赘述。在获取所述用户的身份信息时，可以根据用户的标识信息获取所述身份信息。例如，当用户登录到组织架构的局域网时，可以根据所述用户的标识信息（如登录账号）获取到与所述标识信息相匹配的身份信息，例如，所述标识信息可以

与所述身份信息绑定，在确定用户的标识信息之后，可以根据所述标识信息确定用户的身份信息。当然，在其他实施例中，所述组织架构不限于具有内部局域网，也可以运行于非局域网中，本申请对此不作限制。

[24]本说明书中，用户可以在组织架构内搜索数据，具体地，用户可以在所述组织架构的局域网内输入搜索词，如“卷积神经网络”、“医疗 VR”等等搜索词。在获取到所述搜索词之后，可以获取与所述搜索词相匹配的搜索结果项集合。本说明书中，所述搜索结果项可以包括多种数据类型，如文档、图片、视频、资讯、日志等多种数据信息。所述搜索结果项集合中可以包括上述数据类型中的一种或者多种。其中，所述搜索结果项与所述搜索词相匹配可以包括所述搜索结果项中可以包含所述搜索词，例如，对于文档搜索来说，搜索到的文档中可以包含所述搜索词，如文档的标题、摘要、正文、结语等一处或者多处出现所述搜索词。当然，在说明书的其他实施例中，所述搜索结果项与所述搜索词相匹配还可以包括所述搜索结果项中包括与所述搜索词的同义或者相似含义的词汇，对此，本申请在此不做限制。

[25]S103：获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据。

[26]本说明书中，还可以获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据。本说明书中，可以从所述组织架构的数据库中获取用户的用户日志，并从所述用户日志中获取所述用户对所述搜索结果项集合的历史访问数据。本说明书中，所述与所述用户的所述身份信息相匹配的其他用户可以包括与所述用户的身份信息相同的多个其他用户，或者，与所述用户的身份信息位于同一等级的多个其他用户。例如，若用户 A 属于公司研发部门的研发人员，即与用户 A 相匹配的多个其他用户可以包括与所述用户 A 同样属于研发部门的多个研发人员。本说明书中，所述历史访问数据可以包括所述用户和所述多个其他用户对所述搜索结果项的访问频率、访问时间等多个数据。例如，在一个示例中，根据用户 A 的搜索词“卷积神经网络”搜索到共 200 篇与卷积神经网络相关的电子文档，根据后台数据库的用户日志，可以获取到用户 A 分别对所述 200 篇电子文档的访问记录，如对各篇电子文档的访问次数等数据。另外，在确定与用户 A 相匹配的多个其他用户，即与用户 A 同样属于研发部门的多个研发人员之后，可以获取所述多个研发人员的对上述 200 篇电子文档的访问记录。

[27]S105：根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所

述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[28]本说明书中，在获取到所述用户以及与所述用户身份信息相匹配的多个其他用户的历史访问数据之后，可以根据所述用户以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。在一个实施例中，如图 2 所示，所述根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项可以包括：

[29]S201：根据所述用户以及所述多个其他用户的历史访问数据，分别统计所述用户以及所述多个其他用户分别对所述搜索结果项的访问频率。

[30]S203：根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合。

[31]S205：根据所述用户的历史访问数据以及所述频繁项集合，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[32]本说明书中，可以根据所述用户以及所述多个其他用户的历史访问数据，分别统计所述用户以及所述多个其他用户对所述搜索结果项集合中搜索结果项的访问频率。例如在上述举例中，用户 A 对文档 3 的访问频率为 3，即用户 A 一共访问过 3 次文档 3，而用户 B 对文档 80 的访问频率为 10，即用户 B 一共访问过 10 次文档 80。然后，可以根据所述访问频率从所述搜索结果项中提取所述身份信息所对应的频繁项集合。所述频繁项集合可以包括所述身份信息所对应的访问比较频繁的多个搜索结果项组成的集合。具体地，在一个实施例中，如图 3 所示，所述根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合可以包括：

[33]S301：根据所述访问频率分别统计所述搜索结果项集合中搜索结果项的总访问频率。

[34]S303：根据所述搜索结果项的总访问频率计算所述搜索结果项的支持度值。

[35]S305：根据所述搜索结果项的支持度值计算所述搜索结果项集合的总支持度值。

[36]S307：将所述搜索结果项集合中支持度值不小于所述总支持度值的搜索结果项的集合作为所述身份信息对应的频繁项集合。

[37]本说明书中，可以根据所述用户以及所述多个其他用户分别对所述搜索结果项的访问频率，统计各个搜索结果项的总访问频率。例如，对于上述示例中的文档 3，根据各

个用户对文档 3 的访问频率,可以统计得到用户 A 以及其他多个研发人员对文档 3 的总访问频率,例如,据统计,对文档 3 的总访问频率为 40 次。然后,在获取到所述搜索结果项集合中各个搜索结果项的访问频率之后,可以计算各个搜索结果项的支持度值,所述支持度值可以包括各个搜索结果项的关注热度信息。

- 5 [38]在本说明书的其他实施例中,所述身份信息还可以包括多个身份等级,此时,如图 4 所示,所述根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合可以包括:

[39]S401: 根据所述访问频率分别从所述搜索结果项集合中分别提取所述多个身份等级对应的频繁项集合。

- 10 [40]S403: 将所述多个身份等级对应的频繁项集合的交集作为所述身份信息所对应的频繁项集合。

[41]在本实施例中,所述身份等级可以包括组织架构内垂直的用户等级划分。例如,对于一般的企业来说,可以划分为多个部门,每个部门具有各自的职位等级,如决策层、管理层以及普通用户层。也就是说,对于企业内的一个用户来说,该用户的身份信息中可以包括多个身份等级信息。例如上述示例中用户 A,其身份信息可以包括{研发部门,普通研发人员},即用户 A 的身份信息中至少包含两层身份等级。在获取到用户的身份等级之后,可以根据所述访问频率从所述搜索结果项集合中分别提取所述多个身份等级所对应的频繁项集合。在获取到所述多个身份等级所对应的频繁项集合之后,可以将所述多个身份等级所对应的频繁项集合的交集作为所述身份信息所对应的频繁项集合。

- 20 [42]下面通过一个示例说明上述获取频繁项集合的过程,例如,根据搜索词搜索得到的搜索结果项集合可以表示为 $I = \{i_1, i_2, \dots, i_m\}$, i_m 表示为第 m 个搜索结果项。根据组织架构局域网中的用户日志等历史数据,统计得到组织架构中搜索部门的用户 A 实际访问的点击项集合为 $H_i = \{h_1, h_2, \dots, h_k\}$, h_k 表示为第 k 个点击项,显然 $k \leq m$ 且 $H_i \subset I$ 。对于该搜索部门内的所有 u 个用户,可以得到具有 u 个不同点击项集合的点击项总集合
- 25 $D_t = \{H_1, H_2, \dots, H_u\}$ 。统计所述计算点击项总集合 D_t 中所有的点击项,得到点击项总集合 D_t 中共包含 P 个点击项 $H_t = \{h_1, h_2, \dots, h_p\}$, 其中, h_i 表示 H_t 一项,若设置所述 P 个点击项的支持度为 $S_t = \{s_1, s_2, \dots, s_p\}$, 其中, s_i 表示 S_t 一项,则 s_i 的计算公式可以为:

$$s_i = \frac{h_i \text{项的点击次数}}{u \text{个用户总点击次数}}$$

[43] 其中, $n \leq m$, $0 < s_i < 1$ 且 $H_i \subset I$ 。将 $[0,1]$ 区间划分为 n 个等份, 每个点击项的支持度 S_i 分布在该区间各个子区间段上, 那么求得部门总支持度值 $E(S_i)$ 为

$$E(S_i) = \sum_{i=1}^n s_i \frac{i}{n}$$

[44] 根据求出的总支持度 $E(S_i)$ 可以计算得到推荐给该搜索部门用户的频繁项集合 C_i ,

具体地, 可以将点击项总集合 D_i 中支持度大于等于 $E(S_i)$ 的点击项的集合作为所述频繁

5 项集合 C_i 。假设所述频繁项集合 C_i 具有 j 项, 且 $C_i = \{c_1, c_2, \dots, c_j\}$, 则 C_i 中任意一个搜索结果项的支持度值大于等于 $E(S_i)$ 。

[45] 在本说明书的一个实施例中, 如图 5 所示, 所述从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项可以包括:

[46] S501: 从所述搜索词中提取至少一个搜索分词;

10 [47] S503: 分别计算所述至少一个搜索分词在所述搜索结果项中多个搜索区域的邻近度值;

[48] S505: 根据所述邻近度值以及所述多个搜索区域的权重值, 分别计算所述搜索词对应于所述搜索结果项的邻近度值;

15 [49] S507: 根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[50] 本实施例中, 所述搜索词中可以包括一个或者多个独立的分词, 此时, 可以从所述搜索词中提取至少一个搜索分词, 例如对于描述词“医疗 VR”, 可以从中提取出搜索分词“医疗”、“VR”。在确定所述搜索词中的至少一个搜索分词之后, 可以分别计

20 算所述至少一个搜索分词在所述搜索结果项中多个搜索区域的邻近度值。其中, 所述邻近度值越大, 表示所述搜索分词与所述搜索区域的关联程度越大, 就越有可能符合用户的期望结果。所述搜索结果项的搜索区域可以包括所述搜索结果项的各个组成部分, 例如, 当所述搜索结果项为文档时, 所述搜索区域可以包括文档的标题、摘要、正文、结语等多个不同的组成部分。

[51]在获取到所述邻近度值以及所述多个搜索区域的权重值之后，可以分别计算所述搜索词对应于所述搜索结果项的邻近度值。即将多个搜索分词对应的邻近度值合并成所述搜索词对应于所述搜索结果项的邻近度值。在获取到所述搜索词对应于所述搜索结果项的邻近度值之后，可以根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[52]具体地，在本说明书的一个实施例中，所述根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项可以包括：

10 [53]SS1：根据所述用户的历史访问数据、所述多个其他用户的历史访问数据，分别计算所述用户对应于所述搜索结果项的个性化权重值。

[54]SS3：根据所述搜索词对应于所述搜索结果项的邻近度值以及所述用户对应于所述搜索结果项的个性化权重值，计算所述搜索结果项的评分值。

[55]SS5：将所述搜索结果项集合中评分值大于预设阈值的搜索结果项推荐给所述用户。

15 [56]本实施例中，可以根据所述用户以及所述多个其他用户的历史访问数据，分别计算所述用户对应于各个搜索结果项的个性化权重值。根据所述搜索词对应于所述搜索结果项的邻近度值以及所述用户对应于所述搜索结果项的个性化权重值，计算所述搜索结果项的评分值。然后，将所述搜索结果项集合中评分值大于预设阈值的搜索结果项推荐给所述用户。

20 [57]下面通过一个示例具体说明上述实施例过程，对于用户提交的搜索词 q ，通过中文分词得到搜索分词 $\{t_i\}$ （其中， t_i 表示搜索词 q 中的第 i 个搜索分词），且 $t_i \in q$ ，则在某一搜索区域 f 中搜索分词 t_i 的位置记为 $Pos(t_i, f)$ ，搜索分词 t_{i+1} 的位置记为 $Pos(t_{i+1}, f)$ ，搜索分词 t_i 与 t_{i+1} 的最小距离值记为 $Dis_{\min}(t_i, f)$ ，其邻近度值记为 $prox(t_i, t_{i+1}, f)$ ，则

$$Dis_{\min}(t_i, t_{i+1}, f) = \min |Pos(t_{i+1}, f) - Pos(t_i, f)|$$

$$prox(t_i, t_{i+1}, f) = \frac{1}{Dis_{\min}(t_i, t_{i+1}, f) + 1}$$

[58]搜索结果项 d 中包含多个搜索区域的描述（如标题，内容，路径等），且 $f \in d$ ，搜索区域的权重为 $w(f)$ ，则搜索词 q 对应于所述搜索结果项 d 的邻近度值记为 $Prox(q, d)$ ，

那么,

$$\begin{aligned} \text{Prox}(q, d) &= \sum_{f \in d} \sum_{t \in q} \text{prox}(t_i, t_{i+1}, f) \cdot w(f) \\ &= \sum_{f \in d} \sum_{t \in q} \frac{1}{\min | \text{Pos}(t_{i+1}, f) - \text{Pos}(t_i, f) | + 1} \cdot w(f) \end{aligned}$$

[59]通过搜索结果项 d 的邻近度值 $\text{Prox}(q, d)$ 更新评分结果,可以从用户输入的搜索词入手,利用邻近度值对与用户需求相关的搜索结果产生影响,能够准确返回与搜索词相匹配的搜索结果。另外,从邻近度计算公式来看,算法的复杂度没有增加,保证较快的搜索速度。

5

[60]在按照上述实施例方法,根据组织架构内用户点击项集合 H_i 与其所在部门的权限组的频繁项集 V_n 之间不存在包属的关系,但 $H_i \cap V_n \neq \Phi$,系统初始设置用户点击项与推荐项的权值比为常量 $\theta = 50\%$ ($0 < \theta < 1$),当 $\theta = 0$ 时,忽略用户实际点击效果,搜索引擎展现给用户的是推荐的搜索结果;当 $\theta = 100\%$ 时,忽略系统推荐的搜索结果,仅体现用户点击效果。当 $0 < \theta < 100\%$ 时,融合了用户点击和系统推荐的搜索结果。那么用户推荐点击项的权重为 $W(q, d)$ 可以表示为:

10

$$W(q, d) = \begin{cases} \frac{F(h_j)}{\sum_k F(h_k)} \theta + \frac{F'(h_j)}{\sum_{v_{ij} \in V_n} F'(v_{ij})} (1 - \theta) & (h_j \in H_i, v_{ij} \in V_n, h_j = v_{ij}) \\ \frac{F(h_j)}{\sum_k F(h_k)} & (h_j \in H_i, h_j \notin V_n) \\ \frac{F'(v_{ij})}{\sum_{v_{ij} \in V_n} F'(v_{ij})} & (v_{ij} \notin H_i, v_{ij} \in V_n) \end{cases}$$

[61]其中 $F(h_j)$ 表示用户对 h_j 项的点击量, $F'(h_j)$, $F'(v_{ij})$ 分别表示推荐 h_j , v_{ij} 项的点击量。由于处于组织架构内不同部门、不同权限组的用户关注的内容不同,点击密度随时间、部门差别和权限组有较大变化,搜索呈现的排序结果更符合个性化的特点。而且推荐项的权重可以离线计算,定期更新点击项与推荐项的权重,不影响检索速度。

15

[62]评分排序结果融合是先根据查询内容的邻近度值与关键词权重计算计算结果搜索结果的评分值,然后融合结果搜索结果项的个性化权重得到最终排序结果。具体方法如下:

[63]结果搜索结果项的评分分为两个部分,一是基于布尔和向量空间模型的加权评分

$Scorer(q, d)$ ，计算公式可以表示为

$$Scorer(q, d) = \sum_{t \in q} f_T(t) \cdot f_d^2(t)$$

[64] 其中， $f_T(t)$ 表示查询项 t 在某一搜索结果项 d 中出现的词频， $f_d(t)$ 表示含有查询项 t 的搜索结果项数(或占总搜索结果项的比例)。

[65] 二是加权邻近度评分 $Prox(q, d)$ 。因此，扩展布尔和向量空间模型后的评分公式

5 $ExtendScorer(q, d)$ 为

$$ExtendScorer(q, d) = Scorer(q, d) \cdot Prox(q, d)$$

[66] 根据结果搜索结果项的权重，更新评分数据 $ExtendScorer(q, d)$ 。最终得到的结果项集合 D 已知，点击项的权重 $W(q, d)$ 影响集合 D 搜索结果项的评分，原始项权重值为 1，由公式得到融合后的评分结果为

$$\begin{aligned} MergeScorer(q, d) &= ExtendScorer(q, d) \cdot (1 + W(q, d)) \\ &= \left(\sum_{t \in q} f_T(t) \cdot f_d(t)^2 \right) \cdot Prox(q, d) \cdot (1 + W(q, d)) \end{aligned}$$

[67] 对于集合 D 不包含 $W(q, d)$ 的搜索结果项，其评分不变。这样计算得到的评分，能够
10 体现用户行为特征的同时，也向用户推荐了其它用户关注的内容，很好的改善了用户的体验感，从而实现了用户个性化排序结果的展现。

[68] 本申请提供的数据处理方法，可以根据用户的搜索词匹配到搜索结果项集合，再根据用户对所述搜索结果项集合的历史访问数据，以及与所述用户身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据，从所述搜索结果项集合中提取出推荐
15 给所述用户的至少一个搜索结果项。利用用户以及与用户身份信息相匹配的多个用户的历史数据，对推荐给用户的搜索结果项产生影响，使得推荐给用户的搜索结果项更加符合用户的需求以及用户的身份信息。尤其对于处于组织架构内的用户来说，用户在组织架构内搜索到的相关数据更加符合用户的搜索需求以及用户在组织架构内的身份。

[69] 本申请另一方面还提供一种时间序列中异常点检测装置，图 6 是本申请提供的数据处理装置的一种实施例的模块结构示意图，如图 6 所示，所述装置 60 可以包括：
20

结果项获取单元 61，用于获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；

历史数据获取单元 63, 用于获取所述用户对所述搜索结果项集合的历史访问数据, 以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据;

5 推荐结果项提取单元 65, 用于根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

[70]可选的, 在本申请的一个实施例中, 所述推荐结果项提取单元 65 包括:

访问频率统计单元, 用于根据所述用户以及所述多个其他用户的历史访问数据, 分别统计所述用户以及所述多个其他用户分别对所述搜索结果项的访问频率;

10 频繁项集合提取单元, 用于根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合;

结果项提取单元, 用于根据所述用户的历史访问数据以及所述频繁项集合, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

可选的, 在本申请的一个实施例中, 所述频繁项集合提取单元可以包括:

15 总访问频率统计单元, 用于根据所述访问频率分别统计所述搜索结果项集合中搜索结果项的总访问频率;

支持度计算单元, 用于根据所述搜索结果项的总访问频率计算所述搜索结果项的支持度值;

20 总支持度计算单元, 用于根据所述搜索结果项的支持度值计算所述搜索结果项集合的总支持度值;

频繁项集合提取子单元, 用于将所述搜索结果项集合中支持度值不小于所述总支持度值的搜索结果项的集合作为所述身份信息对应的频繁项集合。

[71]可选的, 在本申请的一个实施例中, 当所述身份信息包括多个身份等级时, 所述频繁项集合提取单元可以包括:

25 多个频繁项集合提取单元, 用于根据所述访问频率分别从所述搜索结果项集合中分别提取所述多个身份等级对应的频繁项集合;

频繁项交集计算单元, 用于将所述多个身份等级对应的频繁项集合的交集作为所述身份信息所对应的频繁项集合。

[72]可选的, 在本申请的一个实施例中, 所述推荐结果项提取单元可以包括:

分词提取单元, 用于从所述搜索词中提取至少一个搜索分词;

第一邻近度值计算单元, 用于分别计算所述至少一个搜索分词在所述搜索结果项中多个搜索区域的邻近度值;

- 5 第二邻近度值计算单元, 用于根据所述邻近度值以及所述多个搜索区域的权重值, 分别计算所述搜索词对应于所述搜索结果项的邻近度值;

结果项提取子单元, 用于根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

- 10 [73]可选的, 在本申请的一个实施例中, 所述结果项提取子单元可以包括:

权重值计算单元, 用于根据所述用户的历史访问数据、所述多个其他用户的历史访问数据, 分别计算所述用户对应于所述搜索结果项的个性化权重值;

评分值计算单元, 用于根据所述搜索词对应于所述搜索结果项的邻近度值以及所述用户对应于所述搜索结果项的个性化权重值, 计算所述搜索结果项的评分值;

- 15 结果项推荐单元, 用于将所述搜索结果项集合中评分值大于预设阈值的搜索结果项推荐给所述用户。

[74]本申请另一方面还提供一种数据处理装置, 图 7 是本申请提供的数据处理装置的一种实施例的模块结构示意图, 如图 7 所示, 所述装置 70 可以包括处理器以及用于存储处理器可执行指令的存储器, 所述处理器执行所述指令时可以实现:

- 20 获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合;

获取所述用户对所述搜索结果项集合的历史访问数据, 以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据;

根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据, 从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

- 25 [75]在 20 世纪 90 年代, 对于一个技术的改进可以很明显地区分是硬件上的改进(例如, 对二极管、晶体管、开关等电路结构的改进)还是软件上的改进(对于方法流程的改进)。然而, 随着技术的发展, 当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路

结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件（Programmable Logic Device, PLD）（例如现场可编程门阵列（Field Programmable Gate Array, FPGA））就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片 PLD 上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器（logic compiler）”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言（Hardware Description Language, HDL），而 HDL 也并非仅有一种，而是有许多种，如 ABEL（Advanced Boolean Expression Language）、AHDL（Altera Hardware Description Language）、Confluence、CUPL（Cornell University Programming Language）、HDCal、JHDL（Java Hardware Description Language）、Lava、Lola、MyHDL、PALASM、RHDH（Ruby Hardware Description Language）等，目前最普遍使用的是 VHDL（Very-High-Speed Integrated Circuit Hardware Description Language）与 Verilog。本领域技术人员也应该清楚，只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中，就可以很容易得到实现该逻辑方法流程的硬件电路。

[76]控制器可以按任何适当的方式实现，例如，控制器可以采取例如微处理器或处理器以及存储可由该（微）处理器执行的计算机可读程序代码（例如软件或固件）的计算机可读介质、逻辑门、开关、专用集成电路（Application Specific Integrated Circuit, ASIC）、可编程逻辑控制器和嵌入微控制器的形式，控制器的例子包括但不限于以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320，存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[77]上述实施例阐明的单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备

的组合。

[78]为了描述的方便，描述以上客户端或服务器时以功能分为各种单元分别描述。当然，在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

5 [79]本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

10 [80]本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[81]这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

20 [82]这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

25 [83]在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[84]内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[85]计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或

技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

[86]还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[87]本领域技术人员应明白，本申请的实施例可提供为方法、系统或计算机程序产品。

因此，本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[88]本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[89]本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[90]以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权利要求书

1.一种数据处理方法，所述方法包括：

获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；

5 获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；

根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

2.根据权利要求 1 所述的方法，其特征在于，所述根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项包括：

根据所述用户以及所述多个其他用户的历史访问数据，分别统计所述用户以及所述多个其他用户分别对所述搜索结果项的访问频率；

15 根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合；

根据所述用户的历史访问数据以及所述频繁项集合，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

3.根据权利要求 2 所述的方法，其特征在于，所述根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合包括：

根据所述访问频率分别统计所述搜索结果项集合中搜索结果项的总访问频率；

根据所述搜索结果项的总访问频率计算所述搜索结果项的支持度值；

根据所述搜索结果项的支持度值计算所述搜索结果项集合的总支持度值；

25 将所述搜索结果项集合中支持度值不小于所述总支持度值的搜索结果项的集合作为所述身份信息对应的频繁项集合。

4.根据权利要求 2 所述的方法，其特征在于，当所述身份信息包括多个身份等级时，所述根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合包括：

30 根据所述访问频率分别从所述搜索结果项集合中分别提取所述多个身份等级对应的频繁项集合；

将所述多个身份等级对应的频繁项集合的交集作为所述身份信息所对应的频繁项集合。

5.根据权利要求1所述的方法，其特征在于，所述从所述搜索结果项集合中提取出

5 推荐给所述用户的至少一个搜索结果项包括：

从所述搜索词中提取至少一个搜索分词；

分别计算所述至少一个搜索分词在所述搜索结果项中多个搜索区域的邻近度值；

根据所述邻近度值以及所述多个搜索区域的权重值，分别计算所述搜索词对应于所述搜索结果项的邻近度值；

10 根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

6.根据权利要求5所述的方法，其特征在于，所述根据所述用户的历史访问数据、

15 所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项包括：

根据所述用户的历史访问数据、所述多个其他用户的历史访问数据，分别计算所述用户对应于所述搜索结果项的个性化权重值；

根据所述搜索词对应于所述搜索结果项的邻近度值以及所述用户对应于所述搜索

20 结果项的个性化权重值，计算所述搜索结果项的评分值；

将所述搜索结果项集合中评分值大于预设阈值的搜索结果项推荐给所述用户。

7.一种数据处理装置，所述装置包括：

结果项获取单元，用于获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜

25 索结果项集合；

历史数据获取单元，用于获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；

推荐结果项提取单元，用于根据所述用户的历史访问数据以及所述多个其他用户的

30 历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

8.根据权利要求7所述的装置,其特征在于,所述推荐结果项提取单元包括:

访问频率统计单元,用于根据所述用户以及所述多个其他用户的历史访问数据,分别统计所述用户以及所述多个其他用户分别对所述搜索结果项的访问频率;

5 频繁项集合提取单元,用于根据所述访问频率从所述搜索结果项集合中提取所述身份信息所对应的频繁项集合;

结果项提取单元,用于根据所述用户的历史访问数据以及所述频繁项集合,从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

10 9.根据权利要求8所述的装置,其特征在于,所述频繁项集合提取单元包括:

总访问频率统计单元,用于根据所述访问频率分别统计所述搜索结果项集合中搜索结果项的总访问频率;

支持度计算单元,用于根据所述搜索结果项的总访问频率计算所述搜索结果项的支持度值;

15 总支持度计算单元,用于根据所述搜索结果项的支持度值计算所述搜索结果项集合的总支持度值;

频繁项集合提取子单元,用于将所述搜索结果项集合中支持度值不小于所述总支持度值的搜索结果项的集合作为所述身份信息对应的频繁项集合。

20 10.根据权利要求8所述的装置,其特征在于,当所述身份信息包括多个身份等级时,所述频繁项集合提取单元包括:

多个频繁项集合提取单元,用于根据所述访问频率分别从所述搜索结果项集合中分别提取所述多个身份等级对应的频繁项集合;

25 频繁项交集计算单元,用于将所述多个身份等级对应的频繁项集合的交集作为所述身份信息所对应的频繁项集合。

11.根据权利要求7所述的装置,其特征在于,所述推荐结果项提取单元包括:

分词提取单元,用于从所述搜索词中提取至少一个搜索分词;

30 第一邻近度值计算单元,用于分别计算所述至少一个搜索分词在所述搜索结果项中多个搜索区域的邻近度值;

第二邻近度值计算单元,用于根据所述邻近度值以及所述多个搜索区域的权重值,

分别计算所述搜索词对应于所述搜索结果项的邻近度值；

结果项提取子单元，用于根据所述用户的历史访问数据、所述多个其他用户的历史访问数据以及所述搜索词对应于所述搜索结果项的邻近度值，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

5

12.根据权利要求 11 所述的装置，其特征在于，所述结果项提取子单元包括：

权重值计算单元，用于根据所述用户的历史访问数据、所述多个其他用户的历史访问数据，分别计算所述用户对应于所述搜索结果项的个性化权重值；

10 评分值计算单元，用于根据所述搜索词对应于所述搜索结果项的邻近度值以及所述用户对应于所述搜索结果项的个性化权重值，计算所述搜索结果项的评分值；

结果项推荐单元，用于将所述搜索结果项集合中评分值大于预设阈值的搜索结果项推荐给所述用户。

15 13.一种数据处理装置，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器执行所述指令时实现：

获取用户的身份信息、搜索词以及与所述搜索词相匹配的搜索结果项集合；

获取所述用户对所述搜索结果项集合的历史访问数据，以及与所述用户的所述身份信息相匹配的多个其他用户对所述搜索结果项集合的历史访问数据；

20 根据所述用户的历史访问数据以及所述多个其他用户的历史访问数据，从所述搜索结果项集合中提取出推荐给所述用户的至少一个搜索结果项。

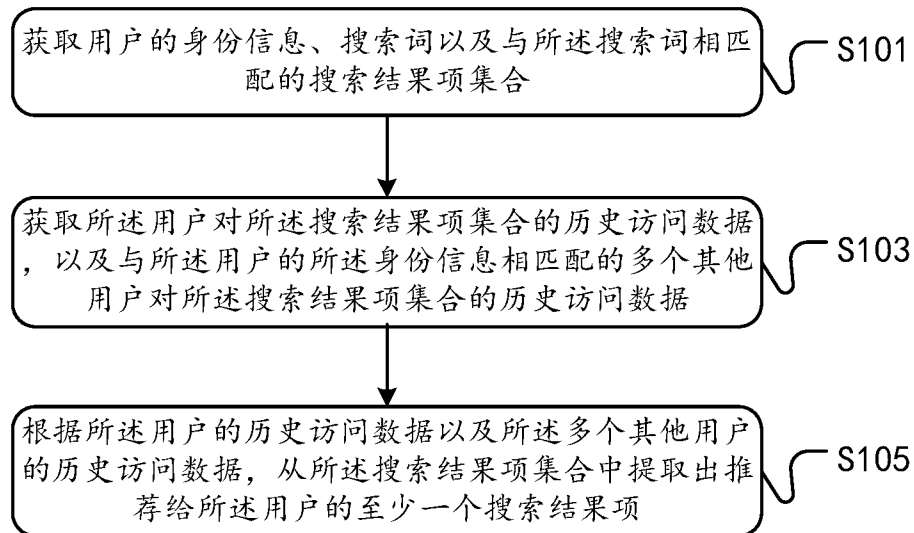


图 1

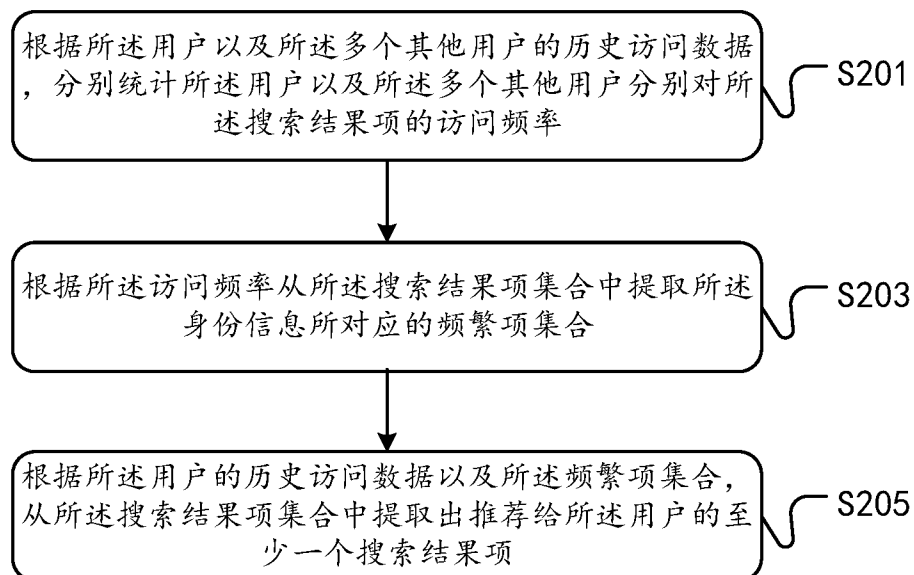


图 2

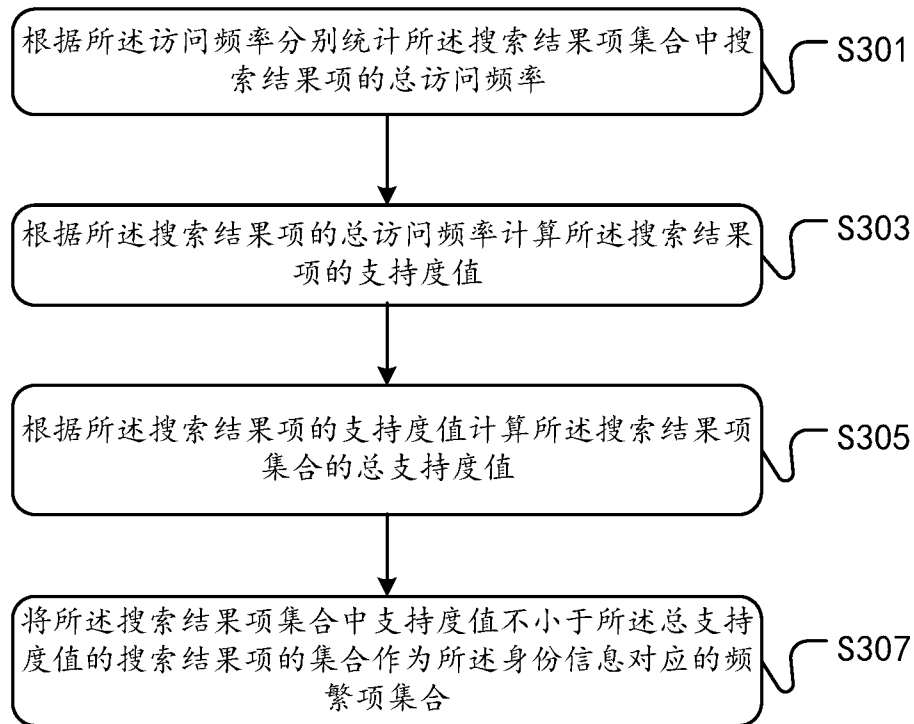


图 3

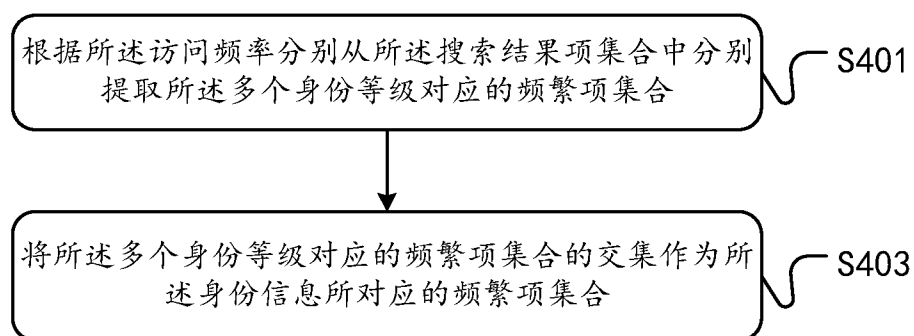


图 4

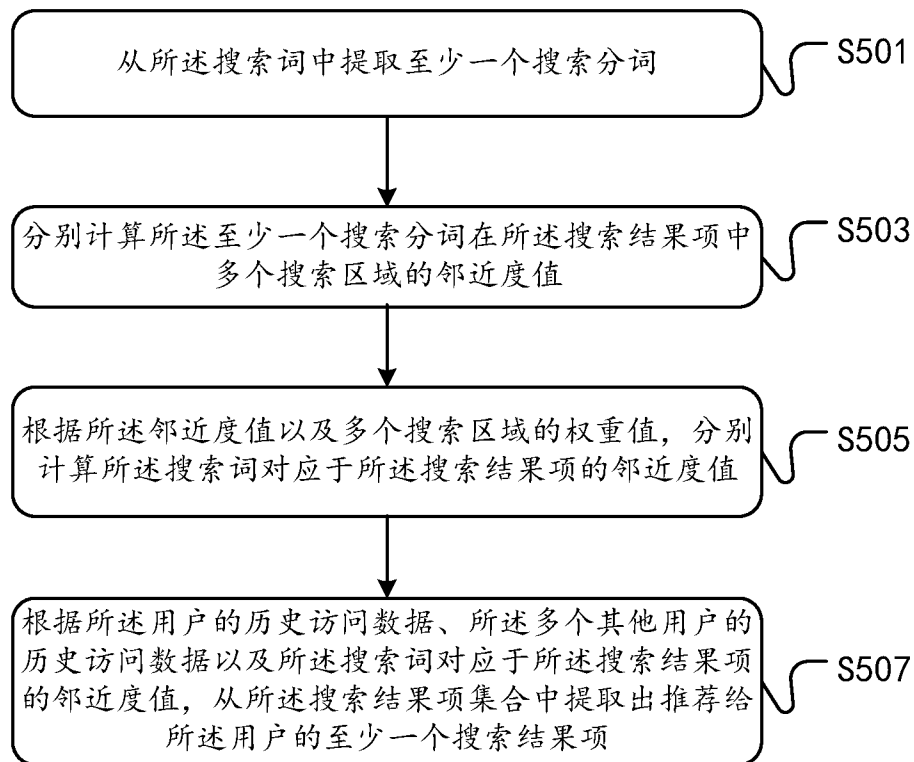


图 5

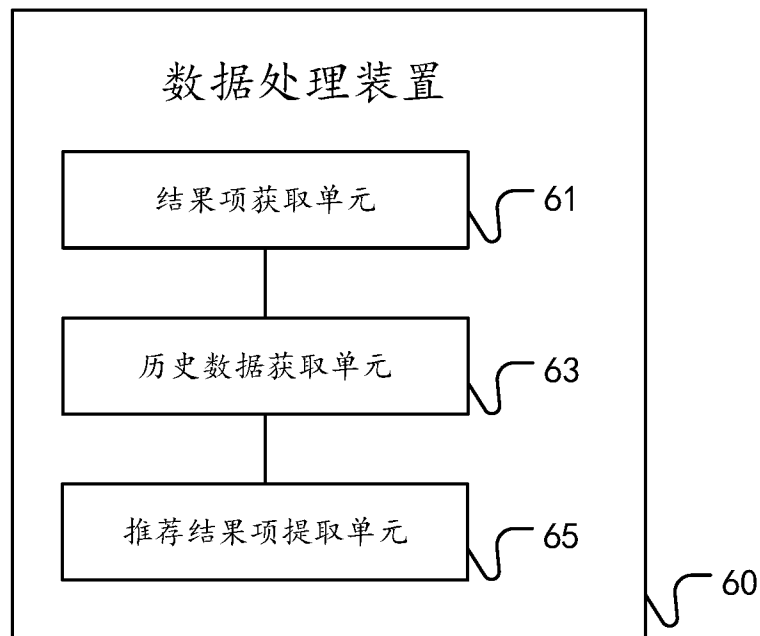


图 6

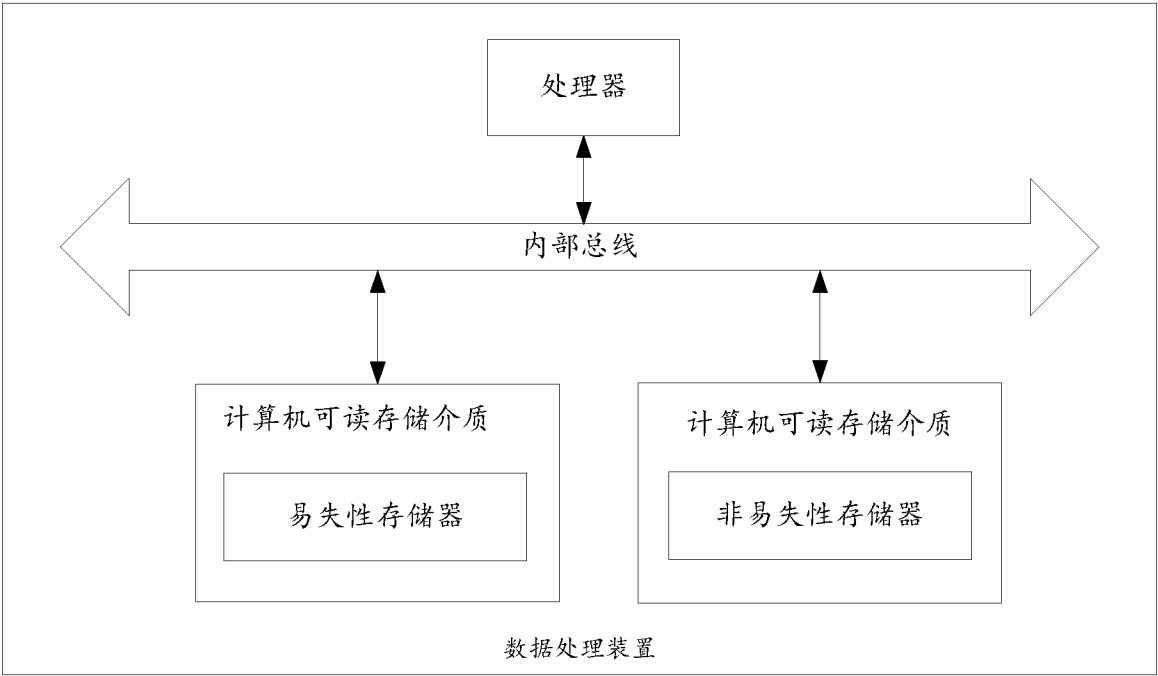


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/099542

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, WPI, EPODOC, CNKI, IEEE: 用户, 身份, 分类, 搜索, 检索, 查询, 查找, 历史, user, ID, classification, search, history

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107908653 A (ALIBABA GROUP HOLDING LIMITED) 13 April 2018 (2018-04-13) claims 1-13	1-13
X	CN 102930022 A (CHINA ACADEMY OF LAUNCH VEHICLE TECHNOLOGY) 13 February 2013 (2013-02-13) description, paragraphs 8-11	1-13
A	CN 103425650 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 04 December 2013 (2013-12-04) entire document	1-13
A	CN 101770474 A (NEWAUTO SILICON VALLEY VIDEO TECHNOLOGY CO., LTD.) 07 July 2010 (2010-07-07) entire document	1-13
A	US 2003004838 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 02 January 2003 (2003-01-02) entire document	1-13
A	US 2004249790 A1 (TOSHIBA TEC KABUSHIKI KAISHA) 09 December 2004 (2004-12-09) entire document	1-13

☐ 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

25 October 2018

Date of mailing of the international search report

14 November 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/099542

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107908653	A	13 April 2018	None			
CN	102930022	A	13 February 2013	None			
CN	103425650	A	04 December 2013	None			
CN	101770474	A	07 July 2010	None			
US	2003004838	A1	02 January 2003	JP	2003030217	A	31 January 2003
US	2004249790	A1	09 December 2004	None			

国际检索报告

国际申请号

PCT/CN2018/099542

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, WPI, EPODOC, CNKI, IEEE: 用户, 身份, 分类, 搜索, 检索, 查询, 查找, 历史, user, ID, classification, search, 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
PX	CN 107908653 A (阿里巴巴集团控股有限公司) 2018年 4月 13日 (2018 - 04 - 13) 权利要求1-13	1-13
X	CN 102930022 A (中国运载火箭技术研究院) 2013年 2月 13日 (2013 - 02 - 13) 说明书第8-11段	1-13
A	CN 103425650 A (腾讯科技深圳有限公司) 2013年 12月 4日 (2013 - 12 - 04) 全文	1-13
A	CN 101770474 A (新奥特硅谷视频技术有限公司) 2010年 7月 7日 (2010 - 07 - 07) 全文	1-13
A	US 2003004838 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2003年 1月 2日 (2003 - 01 - 02) 全文	1-13
A	US 2004249790 A1 (TOSHIBA TEC KABUSHIKI KAISHA) 2004年 12月 9日 (2004 - 12 - 09) 全文	1-13

☐ 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

2018年 10月 25日

国际检索报告邮寄日期

2018年 11月 14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

受权官员

边臻

电话号码 86-(10)-53961419

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/099542

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	107908653	A	2018年 4月 13日	无			
CN	102930022	A	2013年 2月 13日	无			
CN	103425650	A	2013年 12月 4日	无			
CN	101770474	A	2010年 7月 7日	无			
US	2003004838	A1	2003年 1月 2日	JP	2003030217	A	2003年 1月 31日
US	2004249790	A1	2004年 12月 9日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)