

(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
2018 年 12 月 13 日 (13.12.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/223534 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/099711

(22) 国际申请日: 2017 年 8 月 30 日 (30.08.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710433356.8 2017 年 6 月 9 日 (09.06.2017) CN

(71) 申请人: 平安科技 (深圳) 有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(72) 发明人: 余良灵 (YU, Liangling); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。戴从建 (DAI, Congjian); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。方黄威 (FANG, Huangwei); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。叶伟伟 (YE, Weiwei); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。李孝华 (LI, Xiaohua); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一专利商标事务所 (SHENZHEN ZHONGYI PATENT AND TRADEMARK OFFICE);
中国广东省深圳市福田区深南中路

(54) Title: MULTI-SOURCE DATA CATEGORIZATION METHOD AND SERVER

(54) 发明名称: 一种多数据源数据的归类方法及服务器

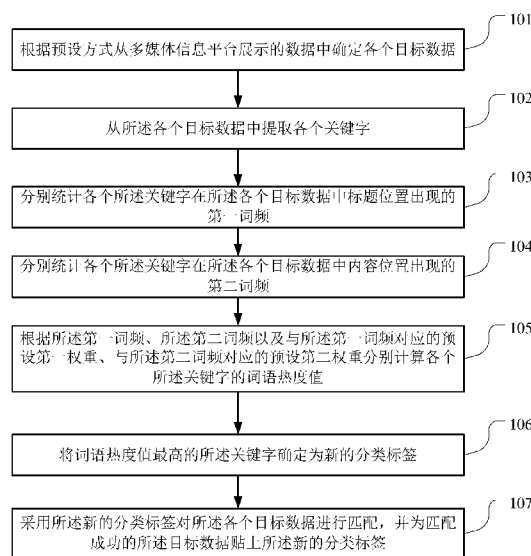


图 1

- 101 Determine target data items according to a preset manner and from data displayed on a multimedia information platform
- 102 Extract keywords from each of the target data items
- 103 Respectively obtain, for each target data item, and for each keyword, a first word frequency for the keyword appearing in a title location
- 104 Respectively obtain, for each target data item, and for each keyword, a second word frequency for the keyword appearing in a content location
- 105 Respectively calculate a word heat index of each of the keywords according to the first word frequency, the second word frequency, a preset first weight corresponding to the first word frequency, and a preset second weight corresponding to the second word frequency
- 106 Determine the keyword having the greatest word heat index as a new category tag
- 107 Perform matching using the new category tag on each target data item and tag successfully matched target data items with the new category tag

(57) Abstract: A multi-source data categorization method and a server, for addressing the problems of inconsistent tag categorization standards and disordered information categorization in a conventional multimedia information platform. The multi-source data categorization method comprises: determining target data items according to a preset manner and from data displayed on a multimedia information platform; extracting keywords from the target data items; respectively obtaining, for each target data item, and for each keyword, a first word frequency for the keyword appearing in a title location and a second word frequency for the keyword appearing in

1014 号老特区报社四楼 (5 号信箱),
Guangdong 518028 (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, 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))。

a content location; respectively calculating a word heat index of each of the keywords according to the first word frequency, the second word frequency, a corresponding first weight, and a corresponding second weight; determining the keyword having the greatest word heat index as a new category tag; and performing matching using the new category tag on each target data item and tagging successfully matched target data items with the new category tag.

(57) 摘要: 一种多数据源数据的归类方法及服务器, 用于解决现有多媒体信息平台分类标签标准不统一, 信息归类混乱的问题。该多数据源数据的归类方法包括: 根据预设方式从多媒体信息平台展示的数据中确定各个目标数据; 从所述各个目标数据中提取各个关键字; 分别统计各个关键字在所述各个目标数据中标题位置出现的第一词频以及在内容位置出现的第二词频; 根据所述第一词频、所述第二词频以及对应的第一权重、对应的第二权重分别计算各个所述关键字的词语热度值; 将词语热度值最高的所述关键字确定为新的分类标签; 采用所述新的分类标签对所述各个目标数据进行匹配, 并为匹配成功的所述目标数据贴上所述新的分类标签。

一种多数据源数据的归类方法及服务器

技术领域

[0001] 本发明涉及数据处理技术领域，尤其涉及一种多数据源数据的归类方法及服务器。

背景技术

[0002] 随着互联网用户对信息的多元化要求越来越高。在构建、维护互联网产品时，需要花费大量的人力、资金来维护产品的多元化信息，以便可以尽可能快地将最新的咨询推送给互联网用户。

[0003] 目前，多媒体信息平台，如音视频平台、文本信息平台等，一般采取整合来自第三方数据源的数据的方式为用户提供多元化的、及时的多媒体信息。在整合多家第三方数据源的数据时，往往需要对这些来自不同地方的数据进行归类，才能在平台上展示时，使得这些数据井然有序。

[0004] 现有的多媒体信息平台一般直接根据第三方数据源提供数据时一并提供的标签对这些数据进行分类。然而，由于多家第三方数据源各自对数据的分类标准不统一，直接采用第三方数据源提供的标签进行分类，往往难以融合多家第三方数据源的数据在同一个平台上，使得平台上展现的多媒体信息产生混乱，难以达到归类的效果。

技术问题

[0005] 本发明实施例提供了一种多数据源数据的归类方法及服务器，能够确定融合多媒体信息平台上的数据的分类标签，为多媒体信息平台对平台上数据的归类整理提供基础，使得多媒体信息平台归类后的数据井然有序，提高归类所追求的效果。

问题的解决方案

技术解决方案

[0006] 第一方面，提供了一种多数据源数据的归类方法，包括：

[0007] 根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

- [0008] 从所述各个目标数据中提取各个关键字；
- [0009] 分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；
- [0010] 分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；
- [0011] 根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；
- [0012] 将词语热度值最高的所述关键字确定为新的分类标签；
- [0013] 采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。
- [0014] 第二方面，提供了一种计算机可读存储介质，所述计算机可读存储介质存储有多数据源数据的归类程序，所述多数据源数据的归类程序被至少一个处理器执行时，实现如下步骤：
- [0015] 根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；
- [0016] 从所述各个目标数据中提取各个关键字；
- [0017] 分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；
- [0018] 分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；
- [0019] 根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；
- [0020] 将词语热度值最高的所述关键字确定为新的分类标签；
- [0021] 采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。
- [0022] 第三方面，提供了一种服务器，所述服务器包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的多数据源数据的归类程序，所述处理器执行所述多数据源数据的归类程序时实现如下步骤：
- [0023] 根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；
- [0024] 从所述各个目标数据中提取各个关键字；
- [0025] 分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；
- [0026] 分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；
- [0027] 根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、

与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；

[0028] 将词语热度值最高的所述关键字确定为新的分类标签；

[0029] 采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。

发明的有益效果

有益效果

[0030] 从以上技术方案可以看出，本发明实施例具有以下优点：

[0031] 本发明实施例中，首先，根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；然后，从所述各个目标数据中提取各个关键字；接着，分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；次之，根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；将词语热度值最高的所述关键字确定为新的分类标签；最后，采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。通过上述步骤，可以根据目标数据中各个关键字的词语热度，从中挑选出词语热度最高的关键字确定为新的分类标签，由于这个新的分类标签是根据多媒体信息平台上的数据确定出来的，其可以融合多媒体信息平台上的数据，不会受到第三方数据源分类标准的影响，从而为多媒体信息平台对平台上数据的归类整理提供了基础，可以使得多媒体信息平台归类后的数据井然有序，提高归类所追求的效果。

对附图的简要说明

附图说明

[0032] 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性的前提下，还可以根据这些附图获得其他的附图。

[0033] 图1为本发明实施例中一种多数据源数据的归类方法第一个实施例流程图；

[0034] 图2为本发明实施例中一种多数据源数据的归类方法第二个实施例流程图；

[0035] 图3是本发明实施例提供的多数据源数据的归类程序的运行环境示意图；

[0036] 图4是本发明实施例提供的多数据源数据的归类程序的功能模块图。

发明实施例

本发明的实施方式

[0037] 本发明实施例提供了一种多数据源数据的归类方法及服务器，用于解决现有多媒体信息平台分类标签标准不统一，信息归类混乱的问题。

[0038] 为使得本发明的发明目的、特征、优点能够更加的明显和易懂，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，下面所描述的实施例仅仅是本发明一部分实施例，而非全部的实施例。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它实施例，都属于本发明保护的范围。

[0039] 请参阅图1，本发明实施例中一种多数据源数据的归类方法第一个实施例包括：

[0040] 101、根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

[0041] 本实施例中所说的目标数据，是指所述归类方法生成新的归类标签所针对或者所指向的数据，这些目标数据是属于多媒体信息平台上展示的数据，其既可以是在多媒体信息平台上已贴上其它标签的数据，也可以是未贴上标签的数据，但必须是多媒体信息平台已从第三方数据源中获取到的数据。可以理解的是，多媒体信息平台从第三方数据源中获取到新的数据（一般为多媒体信息）之后，这些新的数据要么未贴上相应的标签，要么贴上的的是第三方数据源提供数据之前已贴上的标签。因此，这些新的数据对于多媒体信息平台而言，是未经过整理归类的，可以将其确定为目标数据。

[0042] 一般来说，在确定目标数据时，至少可以包括以下两种方式：第一种方式是，将所述多媒体信息平台展示的数据中最近获取的数据确定为所述目标数据，所述最近获取的数据为所述多媒体信息平台在距离当前时间的预设时间长度内从第三方数据源处获取的数据；第二种方式，将所述多媒体信息平台展示的数据中未贴有分类标签的数据确定为所述目标数据。

[0043] 可以理解的是，对于上述第一种方式，即将多媒体信息平台最近一段时间按新获取到的数据确定为目标数据，对于信息归类而言，针对新的数据进行归类和整理相比归类旧数据的意义更大，在当今信息快速喷发的时代，平台的用户也往往更加关心全新的咨询、信息和数据。对于上述第二种方式，也即将未贴有分类标签的数据确定为目标数据，这里需要注意的是，这里所说的“未贴有分类标签”可以是指未贴有多媒体信息平台提供的标签，也可以是指未贴有任何标签（包括第三方数据源提供的标签），在实际应用中，具体选择哪种情境下的“未贴有分类标签”的数据作为目标数据，可以根据具体情况而定。

[0044] 需要说明的是，上述的“各个目标数据”是指步骤101确定出的目标数据往往不止一个，一般来说是多个甚至巨量数目的目标数据，这些目标数据的表现形式可以是文件、数据包等，为便于理解和应用，在实际使用时，一般将一个多媒体文件作为一个目标数据来看待。因此，这里的“各个目标数据”往往包括多个多媒体文件。

[0045] 102、从所述各个目标数据中提取各个关键字；

[0046] 在确定各个目标数据之后，可以理解的是，可以从这些目标数据中提取出一个、两个或者多个关键字。一般来说，可以提取到多个关键字。这里的关键字可以为英文、数字、中文以及其它文字的任意结合，也可以是单个的字母、数字、符号，或者词语、单词，等等。

[0047] 具体地，在提取关键字时，针对不同的目标数据的格式，可以采用不同的方法对这些目标数据进行处理和提取关键字，上述步骤102可以包括：若所述目标数据为音频数据，则对所述音频数据进行音频识别，得到音频文本信息；若所述目标数据为视频数据，则对所述视频数据进行图像识别，得到视频文本信息；若所述目标数据为文本数据，则对所述文本数据进行数据清洗，得到清洗后的文本信息；采用正向最大匹配算法从所述音频文本信息、所述视频文本信息和/或所述清洗后的文本信息中提取得到各个关键字。

[0048] 可以理解的是，针对音频格式的目标数据或者视频格式的目标数据，可以采用音频识别技术或者图像识别技术将这些目标数据转换成文本信息；而对于原本就是本文信息的目标数据，为了提取的方便和高效，可以进行数据清洗。这些

说的数据清洗可以具体包括：对文本信息进行必要的处理，比如大写转小写、全角转半角；检查数据一致性；处理无效值和缺失值；等等。在得到这些文本信息（音频文本信息、所述视频文本信息和/或所述清洗后的文本信息）之后，可以采用正向最大匹配算法从这些文本信息中提取得到各个关键字，具体可以是，对英文分词处理，识别出文本信息中的英文词；对中文数量词处理：用来识别文本信息中的中文数量词，比如"八斤"，"一两"；对中文分词处理：识别出文本信息中的中文词，可以采用了机械分词法，比如按照机械分词法对"公路局"分词的提取结果是"公路局"；等等。关于提取关键字的技术内容，本实施例中不做具体限定。

[0049] 103、分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；

[0050] 104、分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；

[0051] 对于上述步骤103和步骤104，可以理解的是，对于多媒体信息而言，一般来说，标题位置出现的关键字比内容位置（摘要或正文）出现的关键字要更加重要，标题位置的信息也往往是一个多媒体信息或多媒体文件的主要信息和总结性的信息。因此，本实施例中，需要分别考量关键字在标题位置出现的词频和在内容位置出现的词频，分别统计为第一词频和第二词频。

[0052] 105、根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；

[0053] 对于步骤105，由上述内容可知，由于关键字出现在一个目标数据的标题位置和内容位置两种情况，这个关键字对于该目标数据来说在这两种情况下的重要程度是不一样的。一般而言，关键字出现在目标数据的标题位置时，其重要程度较高，而出现在内容位置时，其重要程度相对较低。因此，本实施例对于所述第一词频和第二词频分别设置了对应的第一权重和第二权重，该第一权重和第二权重分别用于表征第一词频和第二词频的重要程度，也即对于计算得出的词语热度值的贡献程度。在实际使用时，该第一权重和第二权重的值可以预先

设定，且可以根据实际情况具体设定这两个值的大小。一般来说，第一权重要高于第二权重，例如可以设定第一权重为0.9，第二权重为0.1。

[0054] 本实施例中，词语热度值用于表征各个关键字之间的相对重要程度，同时也是考量各个关键字在该多媒体信息平台上的热度，某个关键字的词语热度值越大，则表示这个关键字在多媒体信息平台上的热度越高，受到平台用户的关注度越高；反正，某个关键字的词语热度值越小，则表示这个关键字在多媒体信息平台上的热度越低，受到平台用户的关注度也越低。

[0055] 106、将词语热度值最高的所述关键字确定为新的分类标签；

[0056] 本实施例中，可以将词语热度值最高的那个关键字确定为新的分类标签，作为该多媒体信息平台的归类标准。

[0057] 107、采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。

[0058] 在确定出新的分类标签之后，可以采用这个新的分类标签对上述确定出的各个目标数据进行匹配。可以理解的是，匹配的方式也可以根据实际使用情况而定，例如可以将该新的分类标签的关键字与某个目标数据的标题位置的信息进行匹配，若这个目标数据标题位置存在与所述关键字相同的信息时，则认为该目标数据与新的分类标签匹配成功；或者，可以将该新的分类标签的关键字与某个目标数据的全文信息进行匹配，若这个目标数据的全文中存在与所述关键字相同的信息时，则认为该目标数据与新的分类标签匹配成功。

[0059] 在目标数据匹配成功之后，即可以为匹配成功的目标数据贴上新的分类标签，也即完成了这个目标数据的归类，归为“新的分类标签”所属的一类。

[0060] 需要说明的是，对于本实施例中的多媒体信息平台来说，其在进行归类时，除了通过上述步骤107采用新的分类标签进行归类以外，还可以采用多媒体信息平台上已有的其它分类标签进行数据的归类处理，比如，多媒体信息平台以前通过执行本实施例的多数据源数据的归类方法产生的分类标签，或者多媒体信息平台的管理人员认为设定的分类标签，如“新闻”、“科技”、“教育”等标签。另外，多媒体信息平台中的标签还可以由数据的类型来确定，比如“音频”标签、“视频”标签等。特别地，多媒体信息平台上还可以设置有“热点”标签，这个标签可

以由管理人员主动设定，其可以将多媒体信息平台上统计播放次数较多、或者统计的评论数较多、或者统计的收藏数较多的数据归类到这个“热点”标签下，通过“热点”标签对上述举例的这些数据进行归类。

[0061] 本实施例中，首先，根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；然后，从所述各个目标数据中提取各个关键字；接着，分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；次之，根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；将词语热度值最高的所述关键字确定为新的分类标签；最后，采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。通过上述步骤，可以根据目标数据中各个关键字的词语热度，从中挑选出词语热度最高的关键字确定为新的分类标签，由于这个新的分类标签是根据多媒体信息平台上的数据确定出来的，其可以融合多媒体信息平台上的数据，不会受到第三方数据源分类标准的影响，从而为多媒体信息平台对平台上数据的归类整理提供了基础，可以使得多媒体信息平台归类后的数据井然有序，提高归类所追求的效果。

[0062] 为便于理解，下面将从另一方面对本发明中一种多数据源数据的归类方法进行更加详细的描述。请参阅图2，本发明实施例中一种多数据源数据的归类方法第二个实施例包括：

[0063] 201、获取来自第三方数据源的新的数据；

[0064] 第二实施例中，与上述第一实施例不同之处为，可以根据多媒体信息平台获取第三方数据源提供的新的数据之后，再根据需要判断是否确定出新的分类标签进行数据的归类处理。

[0065] 因此，步骤201可以先获取来自第三方数据源的新的数据，该步骤的执行主体可以是上述的多媒体信息平台。

[0066] 202、采用所述已设定的分类标签对获取到的所述新的数据进行匹配；

[0067] 在获取到新的数据之后，可以认为这些新的数据均未贴上多媒体信息平台特有

的标签（也即非第三方数据源提供的标签），因此可以先采用已设定的分类标签对这些新的数据进行匹配。

[0068] 203、若所述新的数据匹配成功，则为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签；

[0069] 当新的数据匹配成功时，则可以为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签。例如，多媒体信息平台上已设定的分类标签包括a、b、c三个，新的数据包括k、i、j三个多媒体文件，在将这些分类标签与这些多媒体文件匹配之后发现，分类标签b与多媒体文件i匹配，则为多媒体文件i贴上分类标签b。

[0070] 204、若所述新的数据未匹配成功，则判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值，若是，则执行步骤205，若否，则执行步骤206；

[0071] 当新的数据未匹配成功时，则表示新的数据与所有的已设定的分类标签均未匹配成功，此时，若要对新的数据进行归类，需要设定新的分类标签。本实施例中，为了严格控制多媒体信息平台上已设定的分类标签的数目，避免多媒体信息平台上的分类标签过多而导致平台上的数据更显凌乱，在设定新的分类标签之前，需要判断当前已有的分类标签的数量是否超过预设的数目阈值。例如，假设数目阈值为10，则表示该多媒体信息平台上最多只能同时存在10个分类标签，当分类标签已满10个时，需要删除一个旧的分类标签，才能设定另外一个新的分类标签。

[0072] 205、从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；

[0073] 当判断得知所述多媒体信息平台上已设定的分类标签的数目已达到预设的数目阈值时，则需要将多媒体信息平台上一个已设定的分类标签删除。可以理解的是，在删除分类标签的选择上，可以选择统计播放量最少的分类标签进行删除，因此一个分类标签下所有数据的统计播放量最少，则表示该分类标签归类的数据对于平台用户来说是相对不受欢迎的，因此该分类标签对多媒体信息平台的重要性就较低，从而可以删除这个分类标签。

[0074] 206、根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

[0075] 上述步骤206与第一个实施例中的步骤101内容大致类似，此处不再赘述两者相同或相似部分的内容。

[0076] 进一步地，本实施例中的步骤206与上述第一个实施例中步骤101的不同之处在于，步骤206还可以将所述播放量最少的分类标签下的所有数据确定为所述目标数据。可以理解的是，对于需要删除或不再使用的分类标签而言，其下的数据将成为没有标签的数据或者至少失去一个该分类标签的数据，因此，对于这些数据来说，可以将这些数据确定为本实施例中的目标数据，以便在后续确定出新的分类标签之后，对这些数据进行匹配、归类。

[0077] 207、从所述各个目标数据中提取各个关键字；

[0078] 上述步骤207与第一个实施例中的步骤102内容大致类似，此处不再赘述。

[0079] 208、分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；

[0080] 209、分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；

[0081] 上述步骤208、209分别与第一个实施例中的步骤103、104内容大致类似，此处不再赘述。

[0082] 210、根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；

[0083] 上述步骤210与第一个实施例中的步骤105内容大致类似，此处不再赘述。

[0084] 211、将词语热度值最高的所述关键字确定为新的分类标签；

[0085] 上述步骤210与第一个实施例中的步骤106内容大致类似，此处不再赘述两者相同或相似部分的内容。

[0086] 进一步地，需要说明的是，由上述内容可知，该多媒体信息平台可以设定有数目阈值，当平台上已设定的分类标签达到这个数目阈值时，在确定出新的分类标签之前，需要删除或不再使用另外一个已设定的分类标签，因此，本实施例中，在步骤211之前，可以先从所述多媒体信息平台上删除所述播放量最少的分

类标签。

[0087] 212、采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。

[0088] 上述步骤212与第一个实施例中的步骤107内容大致类似，此处不再赘述。

[0089] 为便于理解，根据图2所描述的实施例，下面以一个实际应用场景对本发明实施例中的一种多数据源数据的归类方法进行描述：

[0090] 在一个应用场景下，某台小汽车上安装有车载的wifi设备，乘客可以通过该wifi设备接入至多媒体信息平台，乘客接入平台后，可以在该平台上进行音频、视频、文章等多媒体信息的阅览或播放。该多媒体信息平台通过三个第三方数据源提供多媒体数据，分别为数据源A、数据源B和数据源C。

[0091] 某一时刻，多媒体信息平台从数据源B中抓取了一部分新的数据，需要对这些新的数据进行归类。通过多媒体信息平台上已设定有的10个分类标签与这些新的数据匹配后发现，没有一个分类标签与这些数据匹配成功。因此，多媒体信息平台需要确定一个新的分类标签，此时又判断得知，10个分类标签是该平台的标签数量上限，从而多媒体信息平台将其中的分类标签Z删除。然后根据这些新的数据确定出一个词语热度值最高的新的分类标签K，然后采用这个新的分类标签K与所述新的数据以及分类标签Z下的数据进行匹配，并为匹配成功的数据贴上该分类标签K。

[0092] 应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

[0093] 对应于上文实施例所述的多数据源数据的归类方法，图3示出了本发明实施例提供的多数据源数据的归类程序的运行环境示意图，为了便于说明，仅示出了与本发明实施例相关的部分。

[0094] 在本实施例中，所述的多数据源数据的归类程序安装并运行于服务器3中。所述服务器3可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。该服务器3可包括，但不仅限于，一个或多个存储器31（图中仅示出一个）、一个或多个处理器32（图中仅示出一个），上述存储器31和处理器32通过总线33

连接。图3仅示出了具有组件31-33的服务器3，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

[0095] 所述存储器31在一些实施例中可以是所述服务器3的内部存储单元，例如该服务器3的硬盘或内存。所述存储器31在另一些实施例中也可以是所述服务器3的外部存储设备，例如所述服务器3上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。进一步地，所述存储器31还可以既包括所述服务器3的内部存储单元也包括外部存储设备。所述存储器31用于存储安装于所述服务器3的应用软件及各类数据，例如所述多数据源数据的归类程序的程序代码等。所述存储器31还可以用于暂时地存储已经输出或者将要输出的数据。

[0096] 所述处理器32在一些实施例中可以是一中央处理器（Central Processing Unit, CPU），微处理器或其他数据处理芯片，用于运行所述存储器31中存储的程序代码或处理数据，例如执行所述多数据源数据的归类程序等。

[0097] 请参阅图4，是本发明实施例提供的多数据源数据的归类程序的功能模块图。在本实施例中，所述的多数据源数据的归类程序可以被分割成一个或多个模块，所述一个或者多个模块被存储于所述存储器31中，并由一个或多个处理器（本实施例为所述处理器32）所执行，以完成本发明。例如，在图4中，所述的多数据源数据的归类程序可以被分割成目标数据确定模块401、关键字提取模块402、第一词频统计模块403、第二词频统计模块404、词语热度计算模块405、新标签确定模块406、数据匹配模块407。本发明所称的模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述所述多数据源数据的归类程序在所述服务器3中的执行过程。以下描述将具体介绍所述模块401-407的功能。

[0098] 目标数据确定模块401，用于根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

[0099] 关键字提取模块402，用于从所述各个目标数据中提取各个关键字；

[0100] 第一词频统计模块403，用于分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；

- [0101] 第二词频统计模块404，用于分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；
- [0102] 词语热度计算模块405，用于根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；
- [0103] 新标签确定模块406，用于将词语热度值最高的所述关键字确定为新的分类标签；
- [0104] 数据匹配模块407，用于采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。
- [0105] 进一步地，所述目标数据确定模块可以包括：
- [0106] 最近获取数据确定单元，用于将所述多媒体信息平台展示的数据中最近获取的数据确定为所述目标数据，所述最近获取的数据为所述多媒体信息平台在距离当前时间的预设时间长度内从第三方数据源处获取的数据；
- [0107] 和/或
- [0108] 未贴标签数据确定单元，用于将所述多媒体信息平台展示的数据中未贴有分类标签的数据确定为所述目标数据。
- [0109] 进一步地，在根据预设方式从多媒体信息平台展示的数据中确定各个目标数据之前，所述的多数据源数据的归类程序还可以包括：
- [0110] 标签数目判断模块，用于判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；
- [0111] 最少播放标签确定单元，用于若所述标签数目判断模块的判断结果为是，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；
- [0112] 所述目标数据确定模块可以包括：第一标签数据确定单元，用于将所述播放量最少的分类标签下的所有数据确定为所述目标数据；
- [0113] 在将词语热度值最高的所述关键字确定为新的分类标签之前，所述的多数据源数据的归类程序还可以包括：最少播放标签删除单元，用于从所述多媒体信息平台上删除所述播放量最少的分类标签。

- [0114] 进一步地，所述的多数据源数据的归类程序还可以包括：
- [0115] 新数据获取模块，用于获取来自第三方数据源的新的数据；
- [0116] 新数据匹配模块，用于采用所述已设定的分类标签对获取到的所述新的数据进行匹配；
- [0117] 第一贴标签模块，用于若所述新的数据匹配成功，则为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签；
- [0118] 阈值判断模块，用于若所述新的数据未匹配成功，则判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；
- [0119] 第一处理模块，用于若所述阈值判断模块的判断结果为是，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；然后触发所述目标数据确定模块，并且在将词语热度值最高的所述关键字确定为新的分类标签之前，从所述多媒体信息平台上删除所述播放量最少的分类标签；
- [0120] 第二处理模块，用于若所述阈值判断模块的判断结果为否，则触发所述目标数据确定模块。
- [0121] 进一步地，所述关键字提取模块可以包括：
- [0122] 音频文本识别单元，用于若所述目标数据为音频数据，则对所述音频数据进行音频识别，得到音频文本信息；
- [0123] 视频文本识别单元，用于若所述目标数据为视频数据，则对所述视频数据进行图像识别，得到视频文本信息；
- [0124] 数据清洗单元，用于若所述目标数据为文本数据，则对所述文本数据进行数据清洗，得到清洗后的文本信息；
- [0125] 提取单元，用于采用正向最大匹配算法从所述音频文本信息、所述视频文本信息和/或所述清洗后的文本信息中提取得到各个关键字。
- [0126] 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。
- [0127] 以上所述，以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参

照前述实施例对本发明进行了详细的说明，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发明各实施例技术方案的精神和范围。

权利要求书

- [权利要求 1] 一种多数据源数据的归类方法，其特征在于，包括：
- 根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；
- 从所述各个目标数据中提取各个关键字；
- 分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；
- 分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；
- 根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；
- 将词语热度值最高的所述关键字确定为新的分类标签；
- 采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。
- [权利要求 2] 根据权利要求1所述的多数据源数据的归类方法，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：
- 将所述多媒体信息平台展示的数据中最近获取的数据确定为所述目标数据，所述最近获取的数据为所述多媒体信息平台在距离当前时间的预设时间长度内从第三方数据源处获取的数据。
- [权利要求 3] 根据权利要求1所述的多数据源数据的归类方法，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：
- 将所述多媒体信息平台展示的数据中未贴有分类标签的数据确定为所述目标数据。
- [权利要求 4] 根据权利要求1所述的多数据源数据的归类方法，其特征在于，在根据预设方式从多媒体信息平台展示的数据中确定各个目标数据之前，还包括：

判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；

若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；

所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：将所述播放量最少的分类标签下的所有数据确定为所述目标数据；

在将词语热度值最高的所述关键字确定为新的分类标签之前，所述多数据源数据的归类方法还包括：从所述多媒体信息平台上删除所述播放量最少的分类标签。

[权利要求 5]

根据权利要求1所述的多数据源数据的归类方法，其特征在于，所述多数据源数据的归类方法还包括：

获取来自第三方数据源的新的数据；

采用所述已设定的分类标签对获取到的所述新的数据进行匹配；

若所述新的数据匹配成功，则为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签；

若所述新的数据未匹配成功，则判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；

若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；然后执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤，并且在将词语热度值最高的所述关键字确定为新的分类标签之前，从所述多媒体信息平台上删除所述播放量最少的分类标签；

若所述多媒体信息平台上已设定的分类标签的数目未达到预设的数目

阈值，则执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤。

[权利要求 6]

根据权利要求1所述的多数据源数据的归类方法，其特征在于，所述采用所述新的分类标签对所述各个目标数据进行匹配包括：

将所述新的分类标签的关键字与所述各个目标数据的标题位置的信息进行匹配，若目标数据的标题位置上存在与所述关键字相同的信息，则将与所述关键字相同的信息所属的目标数据确定为匹配成功的目标数据；

或

将所述新的分类标签的关键字与所述各个目标数据的全文信息进行匹配，若目标数据的全文信息中存在与所述关键字相同的信息，则将与所述关键字相同的信息所属的目标数据确定为匹配成功的目标数据。

[权利要求 7]

根据权利要求1至6中任一项所述的多数据源数据的归类方法，其特征在于，所述从所述各个目标数据中提取各个关键字包括：

若所述目标数据为音频数据，则对所述音频数据进行音频识别，得到音频文本信息；

若所述目标数据为视频数据，则对所述视频数据进行图像识别，得到视频文本信息；

若所述目标数据为文本数据，则对所述文本数据进行数据清洗，得到清洗后的文本信息；

采用正向最大匹配算法从所述音频文本信息、所述视频文本信息和/或所述清洗后的文本信息中提取得到各个关键字。

[权利要求 8]

一种计算机可读存储介质，所述计算机可读存储介质存储有多数据源数据的归类程序，其特征在于，所述多数据源数据的归类程序被至少一个处理器执行时，实现如下步骤：

根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

从所述各个目标数据中提取各个关键字；

分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一

词频；

分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；

根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；

将词语热度值最高的所述关键字确定为新的分类标签；

采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。

[权利要求 9]

根据权利要求8所述的计算机可读存储介质，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：

将所述多媒体信息平台展示的数据中最近获取的数据确定为所述目标数据，所述最近获取的数据为所述多媒体信息平台在距离当前时间的预设时间长度内从第三方数据源处获取的数据。

[权利要求 10]

根据权利要求8所述的计算机可读存储介质，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：

将所述多媒体信息平台展示的数据中未贴有分类标签的数据确定为所述目标数据。

[权利要求 11]

根据权利要求8所述的计算机可读存储介质，其特征在于，在根据预设方式从多媒体信息平台展示的数据中确定各个目标数据之前，还包括：

判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；

若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；

所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据

包括：将所述播放量最少的分类标签下的所有数据确定为所述目标数据；

在将词语热度值最高的所述关键字确定为新的分类标签之前，所述多数数据源数据的归类方法还包括：从所述多媒体信息平台上删除所述播放量最少的分类标签。

[权利要求 12]

根据权利要求8所述的计算机可读存储介质，其特征在于，所述多数数据源数据的归类程序被至少一个处理器执行时还包括：

获取来自第三方数据源的新的数据；

采用所述已设定的分类标签对获取到的所述新的数据进行匹配；

若所述新的数据匹配成功，则为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签；

若所述新的数据未匹配成功，则判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；

若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；然后执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤，并且在将词语热度值最高的所述关键字确定为新的分类标签之前，从所述多媒体信息平台上删除所述播放量最少的分类标签；

若所述多媒体信息平台上已设定的分类标签的数目未达到预设的数目阈值，则执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤。

[权利要求 13]

根据权利要求8所述的计算机可读存储介质，其特征在于，所述采用所述新的分类标签对所述各个目标数据进行匹配包括：

将所述新的分类标签的关键字与所述各个目标数据的标题位置的信息进行匹配，若目标数据的标题位置上存在与所述关键字相同的信息，则将与所述关键字相同的信息所属的目标数据确定为匹配成功的目标

数据；

或

将所述新的分类标签的关键字与所述各个目标数据的全文信息进行匹配，若目标数据的全文信息中存在与所述关键字相同的信息，则将与所述关键字相同的信息所属的目标数据确定为匹配成功的目标数据。

[权利要求 14]

根据权利要求8至11中任一项所述的计算机可读存储介质，其特征在于，所述从所述各个目标数据中提取各个关键字包括：

若所述目标数据为音频数据，则对所述音频数据进行音频识别，得到音频文本信息；

若所述目标数据为视频数据，则对所述视频数据进行图像识别，得到视频文本信息；

若所述目标数据为文本数据，则对所述文本数据进行数据清洗，得到清洗后的文本信息；

采用正向最大匹配算法从所述音频文本信息、所述视频文本信息和/或所述清洗后的文本信息中提取得到各个关键字。

[权利要求 15]

一种服务器，其特征在于，所述服务器包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的多数据源数据的归类程序，所述处理器执行所述多数据源数据的归类程序时实现如下步骤：

根据预设方式从多媒体信息平台展示的数据中确定各个目标数据；

从所述各个目标数据中提取各个关键字；

分别统计各个所述关键字在所述各个目标数据中标题位置出现的第一词频；

分别统计各个所述关键字在所述各个目标数据中内容位置出现的第二词频；

根据所述第一词频、所述第二词频以及与所述第一词频对应的预设第一权重、与所述第二词频对应的预设第二权重分别计算各个所述关键字的词语热度值；

将词语热度值最高的所述关键字确定为新的分类标签；

采用所述新的分类标签对所述各个目标数据进行匹配，并为匹配成功的所述目标数据贴上所述新的分类标签。

[权利要求 16] 根据权利要求15所述的服务器，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：
将所述多媒体信息平台展示的数据中最近获取的数据确定为所述目标数据，所述最近获取的数据为所述多媒体信息平台在距离当前时间的预设时间长度内从第三方数据源处获取的数据。

[权利要求 17] 根据权利要求15所述的服务器，其特征在于，所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：
将所述多媒体信息平台展示的数据中未贴有分类标签的数据确定为所述目标数据。

[权利要求 18] 根据权利要求15所述的服务器，其特征在于，在根据预设方式从多媒体信息平台展示的数据中确定各个目标数据之前，还包括：
判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；
若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；
所述根据预设方式从多媒体信息平台展示的数据中确定各个目标数据包括：将所述播放量最少的分类标签下的所有数据确定为所述目标数据；
在将词语热度值最高的所述关键字确定为新的分类标签之前，所述处理器执行所述多数据源数据的归类程序时还包括：从所述多媒体信息平台上删除所述播放量最少的分类标签。

[权利要求 19] 根据权利要求15所述的服务器，其特征在于，所述处理器执行所述多数据源数据的归类程序时还包括：
获取来自第三方数据源的新的数据；

采用所述已设定的分类标签对获取到的所述新的数据进行匹配；
若所述新的数据匹配成功，则为匹配成功的所述新的数据贴上与所述新的数据对应匹配的所述已设定的分类标签；
若所述新的数据未匹配成功，则判断所述多媒体信息平台上已设定的分类标签的数目是否达到预设的数目阈值；
若所述多媒体信息平台上已设定的分类标签的数目达到预设的数目阈值，则从所述已设定的分类标签中确定一个播放量最少的分类标签，所述播放量最少的分类标签是指其下所有数据的统计播放量最少的分类标签；然后执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤，并且在将词语热度值最高的所述关键字确定为新的分类标签之前，从所述多媒体信息平台上删除所述播放量最少的分类标签；
若所述多媒体信息平台上已设定的分类标签的数目未达到预设的数目阈值，则执行根据预设方式从多媒体信息平台展示的数据中确定各个目标数据的步骤。

[权利要求 20] 根据权利要求15至17中任一项所述的服务器，其特征在于，所述从所述各个目标数据中提取各个关键字包括：
若所述目标数据为音频数据，则对所述音频数据进行音频识别，得到音频文本信息；
若所述目标数据为视频数据，则对所述视频数据进行图像识别，得到视频文本信息；
若所述目标数据为文本数据，则对所述文本数据进行数据清洗，得到清洗后的文本信息；
采用正向最大匹配算法从所述音频文本信息、所述视频文本信息和/或所述清洗后的文本信息中提取得到各个关键字。

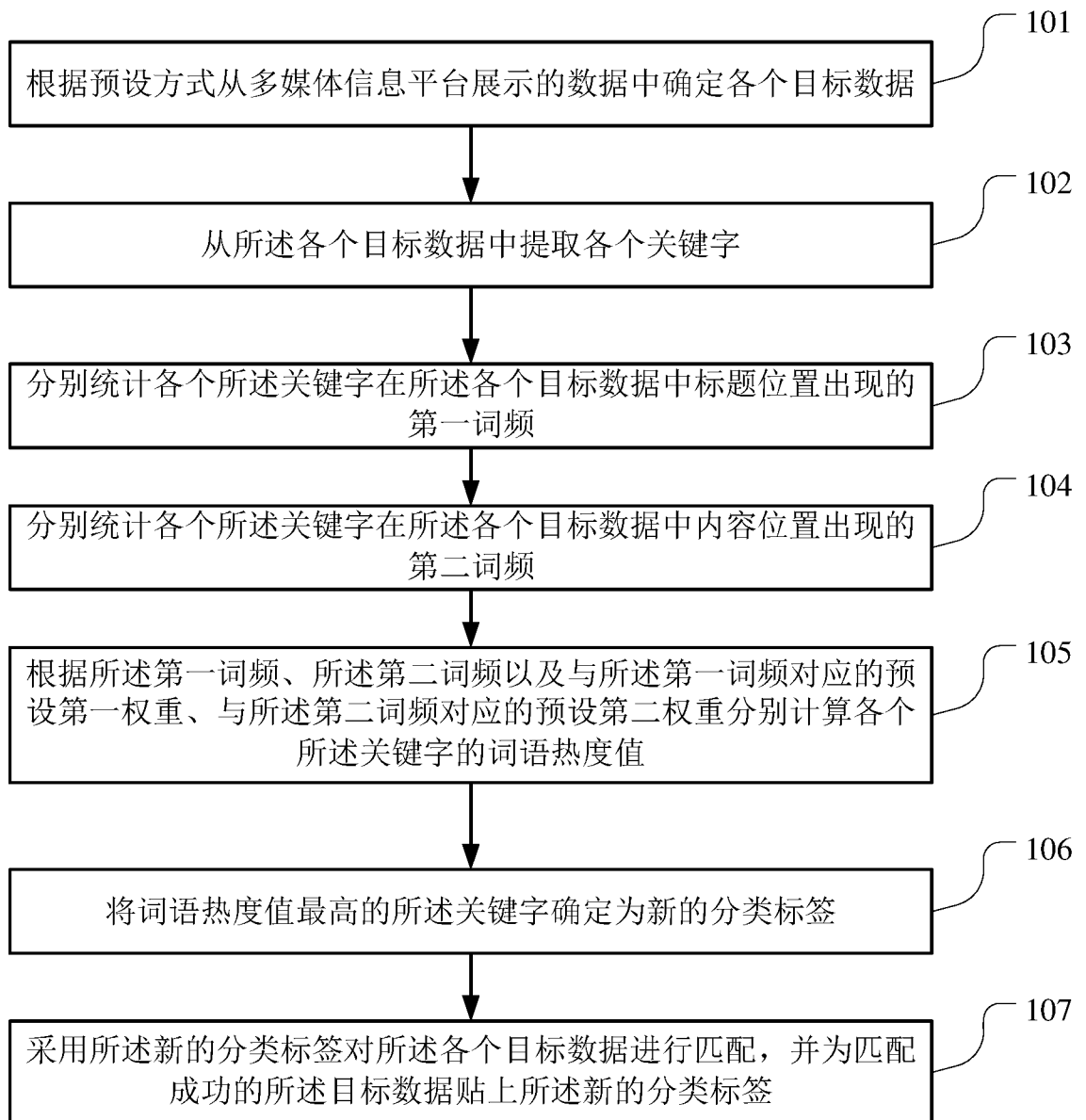


图 1

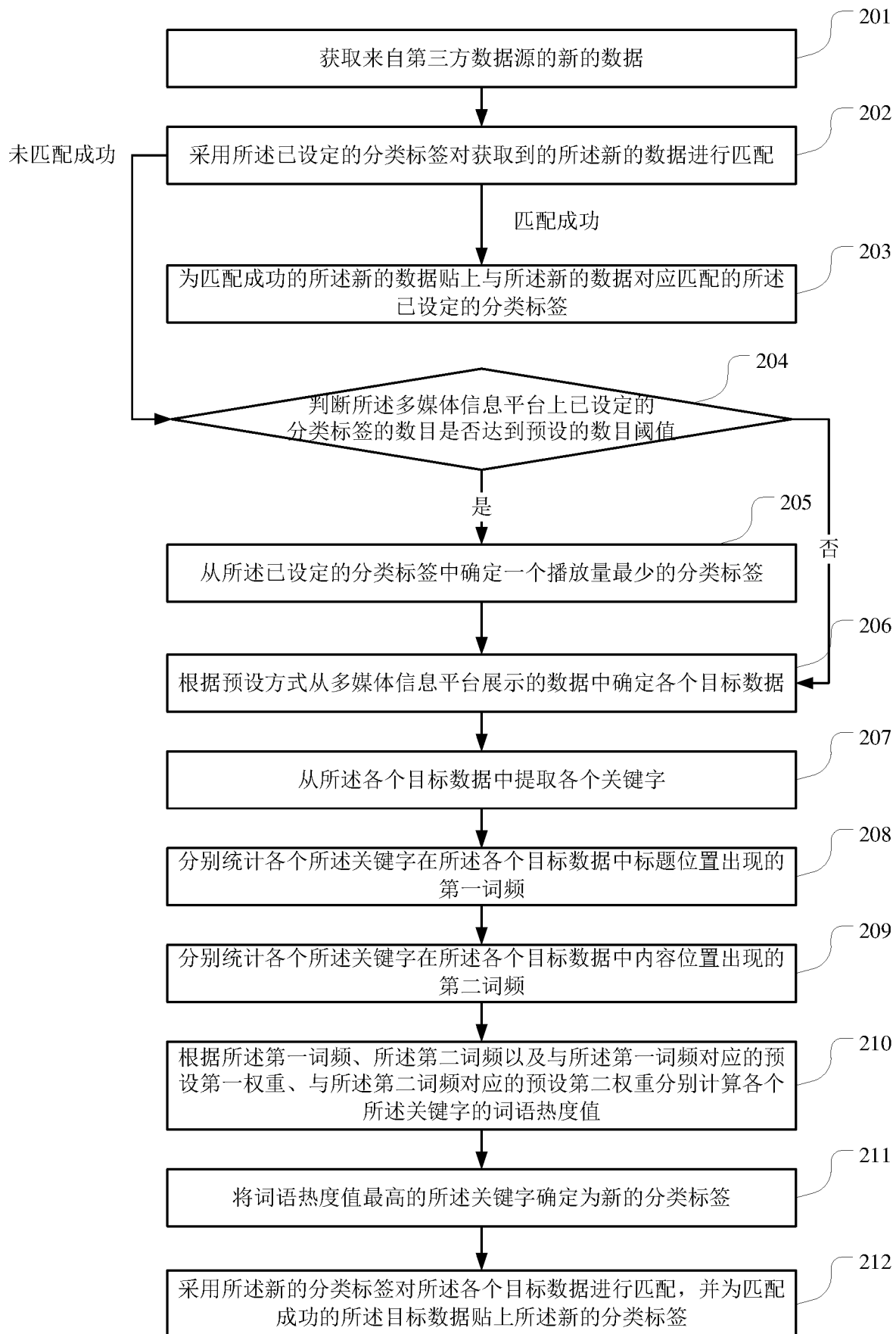


图 2

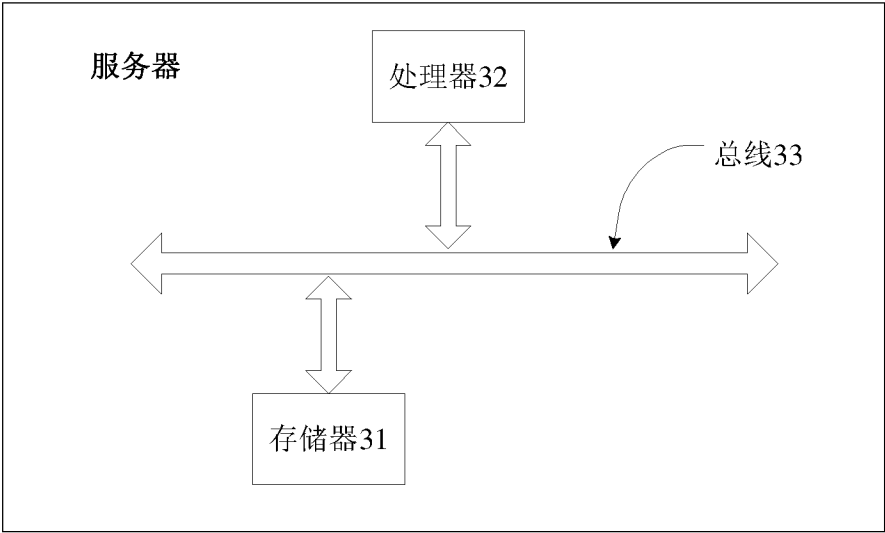


图 3

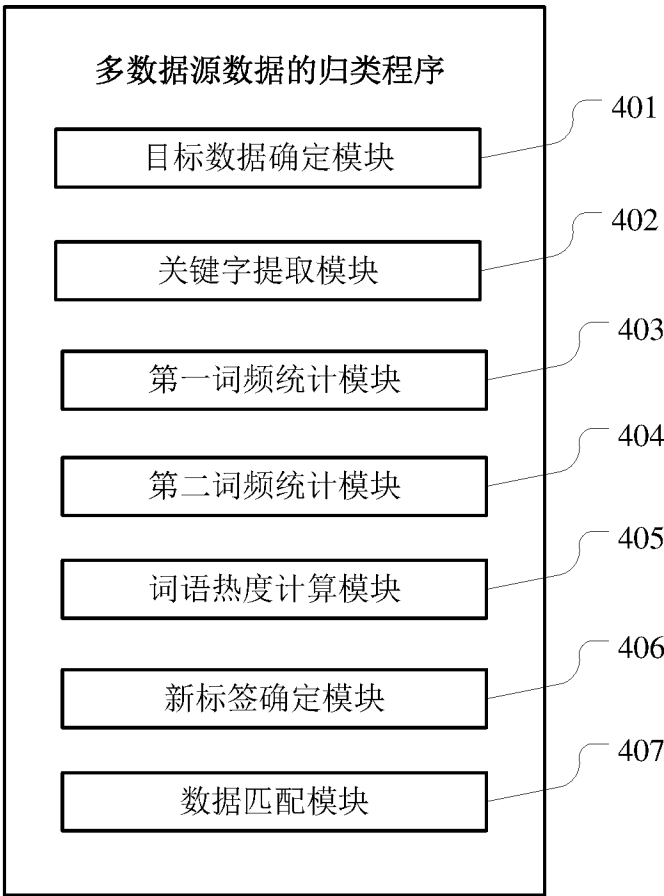


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/099711

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)

CNABS, CNTXT, CNKI, DWPI, SIPOABS: 多, 不同, 源, 数据, 分类, 归类, 关键字, 关键词, 词频, 频率, 权, 标签, multiple, different, data, source?, sort+, class+, keyword, frequency, weight, label

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102831131 A (FUJITSU LIMITED), 19 December 2012 (19.12.2012), description, paragraphs [0031]-[0080] and [0110]-[0138]	1-3, 6-10, 13-17, 20
Y	CN 106528894 A (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.), 22 March 2017 (22.03.2017), description, paragraphs [0074]-[0166]	1-3, 6-10, 13-17, 20
A	CN 104834686 A (INSTITUTE OF INFORMATION ENGINEERING, CHINESE ACADEMY OF SCIENCES), 12 August 2015 (12.08.2015), entire document	1-20

☐ 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 27 February 2018	Date of mailing of the international search report 21 March 2018
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 ZHU, Xiaoli Telephone No. (86-10) 62411672

International application No.
PCT/CN2017/099711

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/099711

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)) CNABS, CNTXT, CNKI, DWPI, SIPOABS: 多, 不同, 源, 数据, 分类, 归类, 关键字, 关键词, 词频, 频率, 权, 标签, multiple, different, data, source?, sort+, class+, keyword, frequency, weight, label														
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 102831131 A (富士通株式会社) 2012年 12月 19日 (2012 - 12 - 19) 说明书第[0031]-[0080]、[0110]-[0138]段</td> <td>1-3, 6-10, 13-17, 20</td> </tr> <tr> <td>Y</td> <td>CN 106528894 A (北京小米移动软件有限公司) 2017年 3月 22日 (2017 - 03 - 22) 说明书第[0074]-[0166]段</td> <td>1-3, 6-10, 13-17, 20</td> </tr> <tr> <td>A</td> <td>CN 104834686 A (中国科学院信息工程研究所) 2015年 8月 12日 (2015 - 08 - 12) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 102831131 A (富士通株式会社) 2012年 12月 19日 (2012 - 12 - 19) 说明书第[0031]-[0080]、[0110]-[0138]段	1-3, 6-10, 13-17, 20	Y	CN 106528894 A (北京小米移动软件有限公司) 2017年 3月 22日 (2017 - 03 - 22) 说明书第[0074]-[0166]段	1-3, 6-10, 13-17, 20	A	CN 104834686 A (中国科学院信息工程研究所) 2015年 8月 12日 (2015 - 08 - 12) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
Y	CN 102831131 A (富士通株式会社) 2012年 12月 19日 (2012 - 12 - 19) 说明书第[0031]-[0080]、[0110]-[0138]段	1-3, 6-10, 13-17, 20												
Y	CN 106528894 A (北京小米移动软件有限公司) 2017年 3月 22日 (2017 - 03 - 22) 说明书第[0074]-[0166]段	1-3, 6-10, 13-17, 20												
A	CN 104834686 A (中国科学院信息工程研究所) 2015年 8月 12日 (2015 - 08 - 12) 全文	1-20												
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。														
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件														
国际检索实际完成的日期 2018年 2月 27日		国际检索报告邮寄日期 2018年 3月 21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 朱晓莉 电话号码 (86-10)62411672												

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/099711

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	102831131	A	2012年 12月 19日	CN 102831131 B	2015年 2月 11日
CN	106528894	A	2017年 3月 22日	无	
CN	104834686	A	2015年 8月 12日	无	