



(12)发明专利申请

(10)申请公布号 CN 111666370 A

(43)申请公布日 2020.09.15

(21)申请号 202010734051.2

(22)申请日 2020.07.28

(71)申请人 中国人民解放军国防科技大学

地址 410073 湖南省长沙市开福区德雅路
109号

(72)发明人 张翀 何春辉 葛斌 刘威
肖建锋

(74)专利代理机构 长沙国科天河知识产权代理
有限公司 43225

代理人 邱轶

(51)Int.Cl.

G06F 16/31(2019.01)

G06F 16/33(2019.01)

G06F 16/9537(2019.01)

G06F 40/30(2020.01)

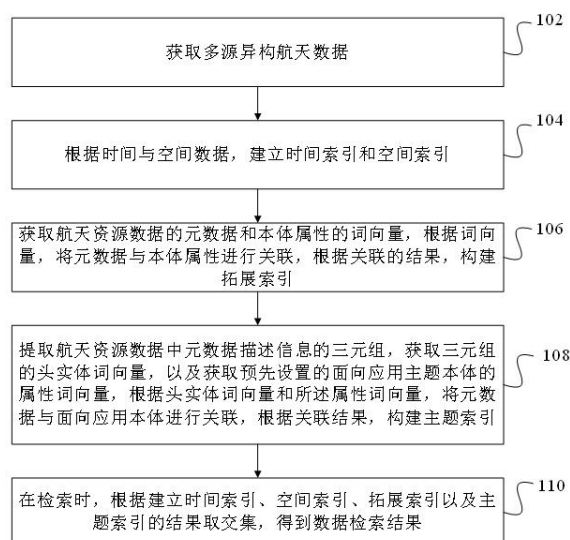
权利要求书2页 说明书10页 附图2页

(54)发明名称

面向多源异构航天数据的语义索引方法和装置

(57)摘要

本申请涉及一种面向多源异构航天数据的语义索引方法和装置。所述方法包括：获取多源异构航天数据；根据时间与空间数据，建立时间索引和空间索引，获取航天资源数据的元数据和本体属性的词向量，根据词向量，将元数据与本体属性进行关联，根据关联的结果，构建拓展索引，提取航天资源数据中元数据描述信息的三元组，获取三元组的头实体词向量，以及获取预先设置的面向应用主题本体的属性词向量，根据头实体词向量和所述属性词向量，将元数据与面向应用本体进行关联，根据关联结果，构建主题索引，在检索时，根据建立时间索引、空间索引、拓展索引以及主题索引的结果取交集，得到数据检索结果。采用本方法满足个性化的检索需求，提高检索效率。



1. 一种面向多源异构航天数据的语义索引方法,其特征在于,所述方法包括:

获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;

获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;

提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

2. 根据权利要求1所述的方法,其特征在于,所述根据时间与空间数据,建立时间索引和空间索引,包括:

提取时间与空间数据的时间值和空间值;

根据所述时间值查询预先设置的时间数组或时间链表对应的时间段,构成时间索引;

将所述空间值插入所述时间段对应的R树,构成空间索引。

3. 根据权利要求2所述的方法,其特征在于,所述方法还包括:

根据所述时间值查询预先设置的时间数组或时间链表,未查询到对应的时间段时,在所述时间数组或时间链表建立所述时间值对应的新增时间段以及所述新增时间段对应的新增R树;

将所述空间值插入所述新增R树中。

4. 根据权利要求1所述的方法,其特征在于,获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,包括:

对所述航天资源数据的元数据和本体属性进行分词处理;所述本体属性是航天信息资源本体的属性;

利用词向量模型在同一空间下训练得到分词处理结果,得到所述航天资源数据的元数据和本体属性的词向量;

根据所述词向量,将所述元数据与所述本体属性进行关联。

5. 根据权利要求4所述的方法,其特征在于,根据所述词向量,将所述元数据与所述本体属性进行关联,包括:

根据所述航天资源数据的元数据的词向量与所述本体属性的词向量的距离,将所述元数据与所述本体属性进行关联。

6. 根据权利要求1所述的方法,其特征在于,提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引,包括:

提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量;

当所述头实体词向量和所述属性词向量的距离小于阈值时,将三元组加入应用主题本体对应的可关联属性列表;否则,将三元组加入非关联属性列表;

根据所述可关联属性列表和所述非关联属性列表建立主题索引。

7. 根据权利要求1至6任一项所述的方法,其特征在于,所述在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果,包括:

在检索时,获取检索条件;

根据所述检索条件、时间索引和空间索引,得到时间与空间数据检索结果;

根据所述检索条件和拓展索引,得到资源本体检索结果;

根据所述检索条件和主题索引,得到应用主题本体检索结果;

取所述时间与空间数据检索结果、资源本体检索结果以及应用主题本体检索结果的交集,得到数据检索结果。

8. 一种面向多源异构航天数据的语义索引装置,其特征在于,所述装置包括:

数据获取模块,用于获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

索引构建模块,用于根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

检索模块,用于在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

9. 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述方法的步骤。

10. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

面向多源异构航天数据的语义索引方法和装置

技术领域

[0001] 本申请涉及计算机技术领域,特别是涉及一种面向多源异构航天数据的语义云索引方法和装置。

背景技术

[0002] 由于航天对地观测任务目的不同,因此观测产生的数据(本发明称航天信息资源)种类不同,如有的是照片格式数据,有的是地理测绘信息数据,并且描述航天信息资源的元数据格式和内容也不同。针对这样的多源异构航天信息资源,用户在面向某些应用需要使用航天信息资源时,往往需要通过检索获得满足条件的全部种类航天信息资源,而不是逐一针对每种航天信息资源进行检索,因此要对各种类航天信息资源构建统一索引。另外,用户在检索时是基于应用的查询,所涉及的属性字段可能原始航天信息资源里没有显式的提供,需要从相关描述信息中构建面向应用的索引。

发明内容

[0003] 基于此,有必要针对上述技术问题,提供一种能够解决多源异构航天数据索引困难的面向多源异构航天数据的语义索引方法和装置。

[0004] 一种面向多源异构航天数据的语义索引方法,所述方法包括:

获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据

根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;

获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;

提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0005] 在其中一个实施例中,还包括:提取时间与空间数据的时间值和空间值;根据所述时间值查询预先设置的时间数组或时间链表对应的时间段,构成时间索引;将所述空间值插入所述时间段对应的R树,构成空间索引。

[0006] 在其中一个实施例中,还包括:根据所述时间值查询预先设置的时间数组或时间链表,未查询到对应的时间段时,在所述时间数组或时间链表建立所述时间值对应的新增时间段以及所述新增时间段对应的新增R树;将所述空间值插入所述新增R树中。

[0007] 在其中一个实施例中,还包括:对所述航天资源数据的元数据和本体属性进行分词处理;所述本体属性是航天资源数据的元数据的应用主题本体的本体属性;利用词向量

模型在同一空间下训练得到分词处理结果,得到所述航天资源数据的元数据和本体属性的词向量;根据所述词向量,将所述元数据与所述本体属性进行关联。

[0008] 在其中一个实施例中,还包括:根据所述航天资源数据的元数据的词向量与所述本体属性的词向量的距离,将所述元数据与所述本体属性进行关联。

[0009] 在其中一个实施例中,还包括:提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量;当所述头实体词向量和所述属性词向量的距离小于阈值时,将三元组加入应用主题本体对应的可关联属性列表;否则,将三元组加入非关联属性列表;根据所述可关联属性列表和所述非关联属性列表建立主题索引。

[0010] 在其中一个实施例中,还包括:在检索时,获取检索条件;根据所述检索条件、时间索引和空间索引,得到时间与空间数据检索结果;根据所述检索条件和拓展索引,得到资源本体检索结果;根据所述检索条件和主题索引,得到应用主题本体检索结果;取所述时间与空间数据检索结果、资源本体检索结果以及应用主题本体检索结果的交集,得到数据检索结果。

[0011] 一种面向多源异构航天数据的语义索引装置,所述装置包括:

数据获取模块,用于获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

索引构建模块,用于根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

检索模块,用于在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0012] 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现以下步骤:

获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;

获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;

提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0013] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:

获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;

获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;

提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0014] 上述面向多源异构航天数据的语义索引方法和装置,针对时间和空间数据,建立时间索引和空间索引,时间索引为数组或链表;空间索引为链接至数组或链表中各个单元的R树,从而实现对时间和空间数据的索引,对于其他航天资源数据,无法直接通过时间索引和空间索引直接检索,通过语意识别的方式,将航天资源数据的元数据映射至航天信息资源本体,从而根据映射关系,建立拓展索引以检索拓展信息,对于面向应用主题的数据,同样采用语意识别的方式,将航天资源数据的元数据映射至应用主题本体,从而建立主题索引,以进行主题检索。通过上述方式,在进行检索时,可以从多个角度检索,而不局限于关键词,从而提高检索的效率。

附图说明

[0015] 图1为一个实施例中面向多源异构航天数据的语义云索引方法的流程示意图;

图2为一个实施例中时间索引和空间索引的示意性结构图;

图3为一个实施例中面向多源异构航天数据的语义云索引装置的结构框图;

图4为一个实施例中计算机设备的内部结构图。

具体实施方式

[0016] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0017] 在一个实施例中,如图1所示,提供了一种面向多源异构航天数据的语义索引方法,包括以下步骤:

步骤102,获取多源异构航天数据。

[0018] 多源异构航天数据包括:时间与空间数据和航天资源数据;时间与空间数据指的是包含时间值和空间值的数据,航天资源数据指的是除时间与空间数据之外的数据。这部分数据的特点是可以通过航天信息资源本体会对航天信息资源中共性的属性做出统一描述和规范,以及存在涉及到相关应用的描述信息。

[0019] 步骤104,根据时间与空间数据,建立时间索引和空间索引。

[0020] 时间索引为数组或链表,如图2所示,空间索引为链接至数组或链表中各个单元的R树。

[0021] 由于时间的连续性,可以将数组或者链表分成多个时间段,该时间段内的数据链接至一个R树,R树中存储空间值,从而在检索时,通过时间与空间数据的时间值,可以检索到对应的时间段,通过空间值,可以检索到在R树中对应的位置。

[0022] 步骤106,获取航天资源数据的元数据和本体属性的词向量,根据词向量,将元数据与本体属性进行关联,根据关联的结果,构建拓展索引。

[0023] 元数据指的是航天资源数据的本身,元数据包含描述信息。

[0024] 本体属性指的是航天信息资源本体的属性,本体有多个,每个本体均被本体属性描述。

[0025] 可以采用语意识别的方式获取元数据和本体属性的词向量,通过词向量,可以确定元数据与本体属性的关联程度,从而将元数据与本体进行关联,通过关联,就可以建立拓展索引。

[0026] 值得说明的是,语意识别以及处理,可以在云端执行,从而在本地无需关注本体与元数据的映射,提高索引建立的效率。

[0027] 步骤108,提取航天资源数据中元数据描述信息的三元组,获取三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据头实体词向量和所述属性词向量,将元数据与面向应用本体进行关联,根据关联结果,构建主题索引。

[0028] 不同的应用产生主题的数据,可以构建不同的主题本体,通过提取元数据描述信息的三元组,可以获取头实体的头实体词向量和应用主题本体的属性词向量,从而将元数据与应用主题本体进行关联,构建主题索引。

[0029] 本步骤的三元组提取和词向量训练也可以在云端执行。

[0030] 步骤110,在检索时,根据建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0031] 检索时,可以进行个性化检索,通过取交集,得到检索结果。

[0032] 上述面向多源异构航天数据的语义索引方法中,针对时间和空间数据,建立时间索引和空间索引,时间索引为数组或链表;空间索引为链接至数组或链表中各个单元的R树,从而实现对时间和空间数据的索引,对于其他航天资源数据,无法直接通过时间索引和空间索引直接检索,通过语意识别的方式,将航天资源数据的元数据映射至航天信息资源本体,从而根据映射关系,建立拓展索引以检索拓展信息,对于面向应用主题的数据,同样采用语意识别的方式,将航天资源数据的元数据映射至应用主题本体,从而建立主题索引,以进行主题检索。通过上述方式,在进行检索时,可以从多个角度检索,而不局限于关键词,从而提高检索的效率。

[0033] 在其中一个实施例中,提取时间与空间数据的时间值和空间值;根据时间值查询预先设置的时间数组或时间链表对应的时间段,构成时间索引;将空间值插入所述时间段对应的R树,构成空间索引。

[0034] 在另一个实施例中,根据时间值查询预先设置的时间数组或时间链表,未查询到对应的时间段时,在时间数组或时间链表建立时间值对应的新增时间段以及新增时间段对

应的新增R树；将空间值插入所述新增R树中。

[0035] 具体的，将时间索引和空间索引统称为骨架索引s-Index，时间索引由数组或链表实现，其中每个元素索引相等时间段 t ，每个元素链接到一个R树的根结点，该R树索引的是该时间段内对应的空间数据，R树的叶子结点中的元素链接对应的时间与空间数据。

[0036] 骨架索引的构架过程具体如下：

S302:取出下一个时间与空间数据 H_i ，将 H_i 的时间值 T_i 和空间值 S_i 提取；

S304:依据 T_i 查找s-Index的时间数组对应的时间段，若无，转入S306，若有，转入S308；

S306:创建对应的时间段元素te，并在te下根据 S_i 创建包含第一个元素 H_i 的R树，返回S302；

S308:设查到的时间段元素为te，向te链接的R树中根据 S_i 值插入 H_i ，返回S302。

[0037] 在其中一个实施例中，对航天资源数据的元数据和本体属性进行分词处理；本体属性是航天资源数据的元数据的应用主题本体的本体属性；利用词向量模型在同一空间下训练得到分词处理结果，得到航天资源数据的元数据和本体属性的词向量；根据词向量，将元数据与本体属性进行关联。

[0038] 在另一个实施例中，根据航天资源数据的元数据的词向量与本体属性的词向量的距离，将元数据与本体属性进行关联。

[0039] 具体的，将元数据与本体属性进行关联的步骤如下：

S402:将航天信息资源本体中的本体属性名 HOA 和本体属性描述 HD ，以及元数据 H 中的属性 A 输入词向量模型word2vec模型中训练；

本体属性描述 HD 需要进行分词处理。

[0040] S404:word2vec将 HO 和 H 在同一空间进行训练后，使得 HOA 和 H 中的 A 都形成词向量；

S406:针对每个航天信息资源种类 C_k 下的每个 A_r^k ，计算与 HO 中的每个与 A_r^k 值类型一致的 HOA_i 词向量的距离 $d_{k,r}^i$ ($0 < d < 1$, 越小越相似)；

S408:挑选最小的距离值 $d_{k,r}^{min}$ ，若 $d_{k,r}^{min}$ 小于等于阈值 σ ，将 A_r^k 关联到 $d_{k,r}^{min}$ 对应的 HOA_{min} 。

[0041] 在关联之后，需要进行索引，具体步骤如下：

S502:针对每个 H_i ，检验其每个 A_j 是否有关联到 HOA ，若有关联转入S504，否则转入S506；

S504:设关联的 HOA 为 HOA_c ，根据 A_j 对应的值 V_j ，将 H_i 插入 HOA_c 对应的索引(HOA_c 为数值型建立B⁺树索引， HOA_c 为文本型建立倒排索引)，返回S502；

S506:根据 A_j 对应的值 V_j ，将 H_i 插入其所在种类 C 的属性 A_j 对应的索引(A_j 为数值型建

立B⁺树索引, A_j 为文本型建立倒排索引), 返回S502。

[0042] 在其中一个实施例中, 提取航天资源数据中元数据描述信息的三元组, 获取三元组的头实体词向量, 以及获取预先设置的面向应用主题本体的属性词向量; 当头实体词向量和属性词向量的距离小于阈值时, 将三元组加入应用主题本体对应的可关联属性列表; 否则, 将三元组加入非关联属性列表; 根据可关联属性列表和非关联属性列表建立主题索引。

[0043] 具体的, 将三元组和应用主题本体进行关联的步骤如下:

S602: 将元数据 H 的值 V 中描述类信息进行三元组抽取, 得到三元组 $\langle s, a, o \rangle$;

S604: 将应用主题本体 TO 中的本体属性名称 TOA 和本体属性描述 TD , 以及所提取的三元组中的头实体 s 输入 word2vec 模型中训练, 其中本体属性描述 TD 要进行分词;

S606: word2vec 将 TO 和 s 在同一空间进行训练后, 使得 TOA 和 s 都形成词向量;

S608: 针对每个 s_j , 计算与 TO 中的每个与 s_j 值类型一致的 TOA_i 词向量的距离 d_j^i ($0 < d < 1$, 越小越相似);

S610: 挑选最小的距离值 d_j^{min} , 若 d_j^{min} 小于等于阈值 ϵ , 转入S6, 否则转入S7;

S612: 将 s_j 关联到 d_j^{min} 对应的 TOA_{min} , 并将 s_j 加入到可关联属性列表 LA ;

S614: 将 s_j 加入到非关联属性列表 NLA 。

[0044] 构建主题索引的步骤包括:

S702: 针对每个 H_i , 将 H_i 的 V_i 中描述类信息进行三元组抽取, 得到 $\langle s, a, o \rangle$;

S704: 检验每个 s_j , 若在 LA 中 (即可关联), 转入S706, 若在 NLA 中 (即不关联), 转入S704, 否则转入S708, 检验完毕后转入S702;

S706: 设关联的 TOA 为 TOA_c , 根据 s_j 对应的值 o_j , 将 H_i 插入 TOA_c 对应的索引 (TOA_c 为数值型建立B⁺树索引, TOA_c 为文本型建立倒排索引), 返回S704;

S708: 再次进行关联, 确定 s_j 是否可关联后转入S704。

[0045] 通过上述方式, 可以构建三种索引, 分别为骨架索引、拓展索引和主题索引。

[0046] 具体在检索时, 获取检索条件; 根据检索条件、时间索引和空间索引, 得到时间与空间数据检索结果; 根据检索条件和拓展索引, 得到资源本体检索结果; 根据检索条件和主题索引, 得到应用主题本体检索结果; 取时间与空间数据检索结果、资源本体检索结果以及应用主题本体检索结果的交集, 得到数据检索结果。

[0047] 具体的, 检索过程如下:

S802: 针对用户提交的时间和空间检索条件, 利用骨架索引 s-Index 过滤时间条件后定位到相关的R树, 再利用R树过滤空间条件, 得到结果 $sret$;

S804:针对用户提交的面向**HOA**的检索条件,利用关联到**HO**的拓展索引e-index进行检索,得到结果**erret**;

S806:针对用户提交的面向航天信息资源本体且没有与**HO**关联的属性检索条件,利用非关联到**HO**的拓展索引e-index进行检索,得到结果**enret**;

S808:针对用户提交的面向**TOA**的检索条件,利用主题索引t-index进行检索,得到结果**tret**;

S810:进行**sret \cap erret \cap enret \cap tret**,得到的结果即为最后检索结果。

[0048] 综上所述,本发明设计了3种索引类型:骨架索引、扩展索引和主题索引,分别用于索引航天信息资源共性属性、航天信息资源个性属性和应用主题属性。从信息输入输出角度看,输入为:

1) 航天信息资源元数据(描述信息)集合 $\{H_1, H_2, \dots, H_n\}$,其中

$H_i = \langle C, url, \{ \langle A_1, V_1 \rangle, \langle A_2, V_2 \rangle, \dots, \langle A_m, V_m \rangle \} \rangle$,其中**C**代表航天信息资源种类,**url**代表此元数据**H_i**所描述的航天信息资源的存储路径,**A**代表属性,**V**代表值,**V**既可以是单值也可以是多值,既可以是数值型数据也可以是文本型数据。

[0049] 2) 航天信息资源本体

$HO = \{ \langle HOA_1, HD_1 \rangle, \langle HOA_2, HD_2 \rangle, \dots, \langle HOA_s, HD_s \rangle \}$,其中**HOA**是属性名,**HD**是对应属性的描述。

[0050] 3) 应用主题本体

$TO = \{ \langle TC, \{ \langle TOA_1, TD_1 \rangle, \langle TOA_2, TD_2 \rangle, \dots, \langle TOA_r, TD_r \rangle \} \rangle \}$,其中**TC**代表应用主题本体的类名,**TOA**是相应类的属性名,**TD**是对应属性的描述。

[0051] 输出为:

1) 以时间和空间信息构建的骨架索引s-Index,骨架索引只有1套。

[0052] 2) 除时间和空间外的关联**HO**属性和未关联**HO**属性的信息构建成的扩展索引e-Index,扩展索引具有多套。

[0053] 3) 面向**TC**类应用关联**TO**属性的信息构建成的主题索引t-Index,主题索引也有多套。

[0054] 4) 基于多源异构航天大数据语义云索引的检索过程。

[0055] 通过上述技术手段,本发明主要解决了如下两个技术问题:

1、航天信息资源多源异构,针对用户检索时统一检索的需求,如何建立能够统领各类航天信息资源的索引。目前很多发明都是依靠模板映射,这在实际中对用户造成较大压力,本发明强调如何能够自动进行映射匹配从而建立统一索引。

[0056] 2、针对用户面向某一主题的应用,如何能够提供给用户超越当前航天信息资源属性字段的检索能力,换言之,用户在针对某一应用A时,不关心航天信息资源的字段,而是面向应用A的属性字段,这如何利用航天信息资源进行建立面向应用主题的索引。

[0057] 应该理解的是,虽然图1和3的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图1和3中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些子步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0058] 在一个实施例中,如图3所示,提供了一种面向多源异构航天数据的语义索引装置,包括:数据获取模块902、索引构建模块904和检索模块906,其中:

数据获取模块902,用于获取多源异构航天数据;所述多源异构航天数据包括:时间与空间数据和航天资源数据;

索引构建模块904,用于根据时间与空间数据,建立时间索引和空间索引;其中,所述时间索引为数组或链表;所述空间索引为链接至所述数组或链表中各个单元的R树;获取所述航天资源数据的元数据和本体属性的词向量,根据所述词向量,将所述元数据与所述本体属性进行关联,根据关联的结果,构建拓展索引;提取所述航天资源数据中元数据描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量,根据所述头实体词向量和所述属性词向量,将所述元数据与所述面向应用本体进行关联,根据关联结果,构建主题索引;

检索模块906,用于在检索时,根据所述建立时间索引、空间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0059] 在其中一个实施例中,索引构建模块904还用于提取时间与空间数据的时间值和空间值;根据所述时间值查询预先设置的时间数组或时间链表对应的时间段,构成时间索引;将所述空间值插入所述时间段对应的R树,构成空间索引。

[0060] 在其中一个实施例中,索引构建模块904还用于根据所述时间值查询预先设置的时间数组或时间链表,未查询到对应的时间段时,在所述时间数组或时间链表建立所述时间值对应的新增时间段以及所述新增时间段对应的新增R树;将所述空间值插入所述新增R树中。

[0061] 在其中一个实施例中,索引构建模块904还用于对所述航天资源数据的元数据和本体属性进行分词处理;所述本体属性是航天信息资源本体的本体属性;利用词向量模型在同一空间下训练得到分词处理结果,得到所述航天资源数据的元数据和本体属性的词向量;根据所述词向量,将所述元数据与所述本体属性进行关联。

[0062] 在其中一个实施例中,索引构建模块904还用于根据所述航天资源数据的元数据的词向量与所述本体属性的词向量的距离,将所述元数据与所述本体属性进行关联。

[0063] 在其中一个实施例中,索引构建模块904还用于提取所述航天资源数据的元数据的描述信息的三元组,获取所述三元组的头实体词向量,以及获取预先设置的面向应用主题本体的属性词向量;当所述头实体词向量和所述属性词向量的距离小于阈值时,将三元组加入应用主题本体对应的可关联属性列表;否则,将三元组加入非关联属性列表;根据所述可关联属性列表和所述非关联属性列表建立主题索引。

[0064] 在其中一个实施例中,检索模块906还用于在检索时,根据所述建立时间索引、空

间索引、拓展索引以及主题索引的结果取交集,得到数据检索结果。

[0065] 关于面向多源异构航天数据的语义索引装置的具体限定可以参见上文中对于面向多源异构航天数据的语义索引方法的限定,在此不再赘述。上述面向多源异构航天数据的语义索引装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0066] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是终端,其内部结构图可以如图4所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口、显示屏和输入装置。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种面向多源异构航天数据的语义索引方法。该计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏,该计算机设备的输入装置可以是显示屏上覆盖的触摸层,也可以是计算机设备外壳上设置的按键、轨迹球或触控板,还可以是外接的键盘、触控板或鼠标等。

[0067] 本领域技术人员可以理解,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0068] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,该存储器存储有计算机程序,该处理器执行计算机程序时实现上述实施例中方法的步骤。

[0069] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现上述实施例中方法的步骤。

[0070] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink) DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0071] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0072] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护

范围。因此,本申请专利的保护范围应以所附权利要求为准。

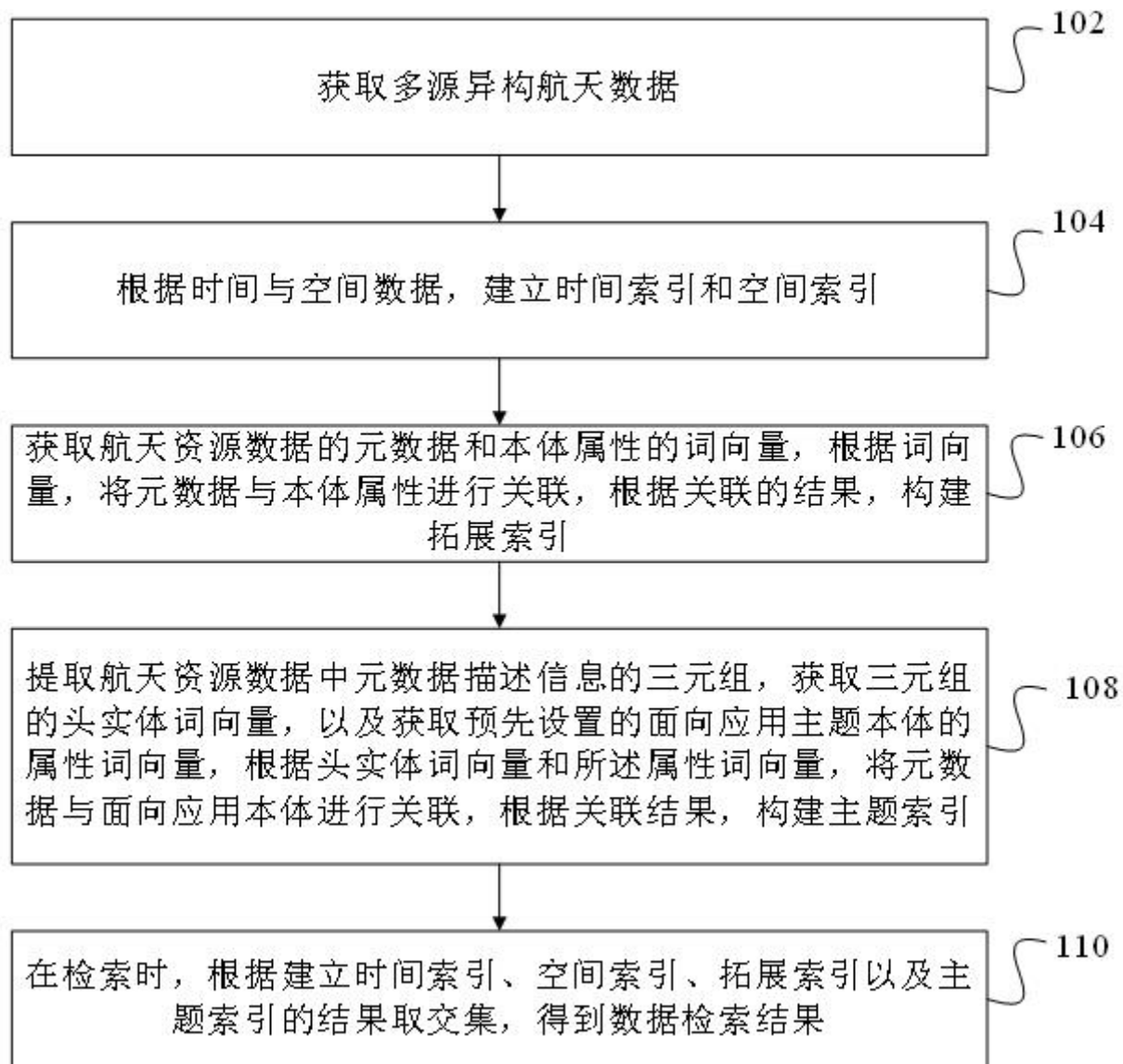


图1

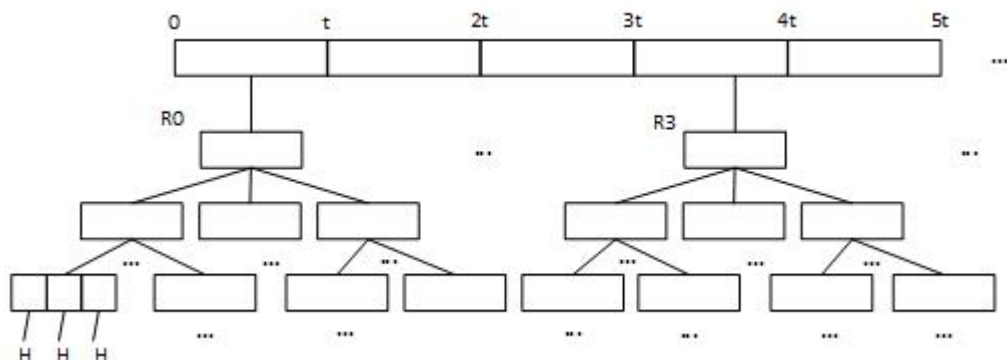


图2

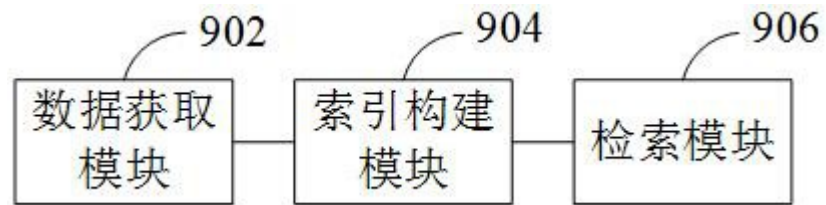


图3

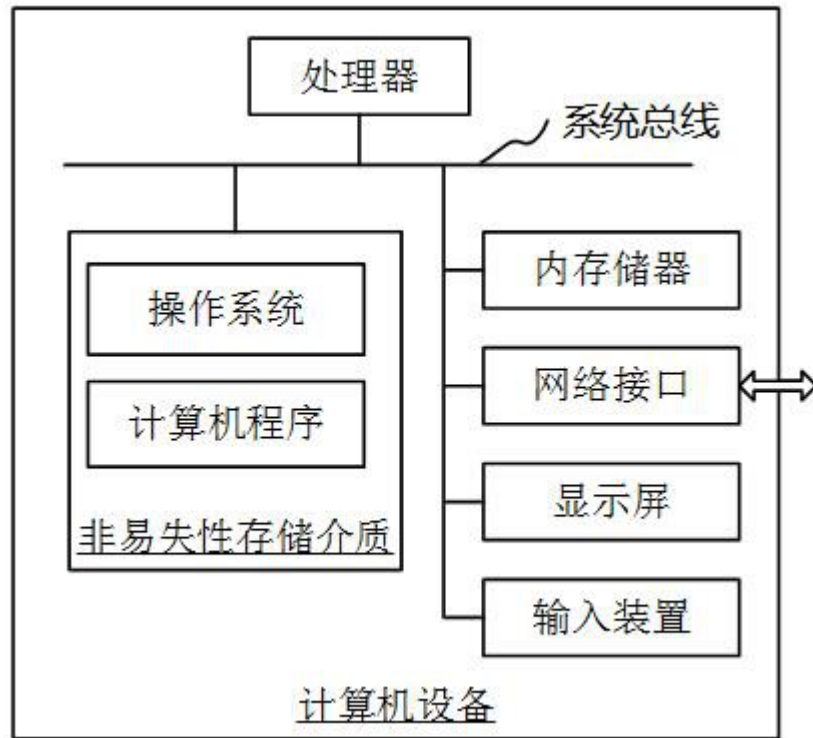


图4