



(12) 发明专利申请

(10) 申请公布号 CN 102737057 A

(43) 申请公布日 2012. 10. 17

(21) 申请号 201110093430. 9

(22) 申请日 2011. 04. 14

(71) 申请人 阿里巴巴集团控股有限公司  
地址 英属开曼群岛大开曼岛资本大厦一座  
四层 847 号邮箱

(72) 发明人 苗文彦 钱建平

(74) 专利代理机构 北京同达信恒知识产权代理  
有限公司 11291  
代理人 郭润湘

(51) Int. Cl.  
G06F 17/30 (2006. 01)

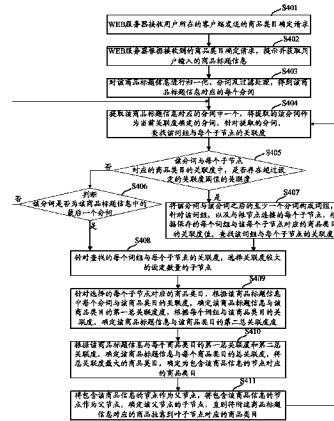
权利要求书 2 页 说明书 12 页 附图 5 页

(54) 发明名称

一种商品类目信息的确定方法及装置

(57) 摘要

本申请公开了一种商品类目信息的确定方法及装置,用以解决现有确定商品类目效率低,准确性差的问题。该方法当获取到商品标题信息后,对该商品标题信息分词构成相应的词组,针对每个词组及与根节点连接的每个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度,并确定包含该商品信息的节点;将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。如本申请提出的方案,因此可以大大缩短确定商品挂靠的叶子节点的时间,保证商品挂靠的商品类目确定的准确性。



1. 一种商品信息与商品类目关联度的确定方法,其特征在于,包括:

针对商品类目树中每个商品类目对应的节点,确定该节点的父节点;

针对该节点包含的每个商品标题信息,进行以下步骤:

对该商品标题信息分词,确定该商品标题信息对应的每个词组,其中所述词组包括至少一个分词;

针对每个词组,根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值;

针对该节点分别保存每个词组与该节点对应的商品类目的关联度值。

2. 如权利要求 1 所述的方法,其特征在于,所述确定该词组与该节点对应的商品类目的关联度值包括:

根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,及该节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的第一权重值;

根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该父节点对应的商品类目中其他商品类目的第二权重值;

确定第一权重值与第二权重值的差值,根据该差值与第一权重值的商,确定该词组与该节点对应的商品类目的关联度值。

3. 如权利要求 2 所述的方法,其特征在于,所述确定该词组与该节点对应的商品类目的第一权重值,包括:

根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,及该节点对应的商品类目包含商品数量的商值,确定第一权重值。

4. 如权利要求 2 所述的方法,其特征在于,所述确定该词组与该父节点对应的商品类目中其他商品类目的第二权重值,包括:

确定该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,及该词组在该节点对应的商品类目包含的商品标题信息中出现的次数的第一差值;

确定该父节点对应的商品类目包含的商品数量,及该节点对应的商品类目包含的商品数量的第二差值;

根据第一差值与第二差值的商值,确定该第二权重值。

5. 一种基于权利要求 1 的关联度的确定方法确定商品类目信息的方法,其特征在于,该方法包括:

接收商品标题信息;

对该商品标题信息进行分词,确定该商品标题信息对应的每个词组,其中每个词组中包括至少一个分词;

根据针对每个节点保存的每个词组与该节点对应的商品类目的关联度值,查找确定的每个词组与每个子节点的关联度;

根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点;

将确定的该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品信息挂靠到叶子节点对应的商品类目。

6. 如权利要求 5 所述的方法,其特征在于,当该词组中包含一个分词时,所述查找每个词组与每个子节点的关联度包括:

针对每个词组,根据保存的每个词组与每个商品类目的关联度值,确定包含该词组的每个商品类目对应的子节点;

针对确定的每个子节点,将保存的该词组与该子节点对应的商品类目的关联度值,作为查找到的该词组与该子节点的关联度。

7. 如权利要求 6 所述的方法,其特征在于,所述方法还包括:

判断查找到的该词组与该子节点的关联度是否大于设定的关联度阈值;

当判断结果为是时,将所述分词与该商品标题信息中的其他分词组成词组,查找该词组与每个商品类目的关联度。

8. 如权利要求 5 所述的方法,其特征在于,根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点,包括:

针对每个词组,根据查找到的该词组与每个子节点的关联度,选择关联度较大的设定数量的子节点;

针对选择的每个子节点对应的商品类目,根据该商品标题信息中每个词组与该商品类目的关联度,确定该商品标题信息与该商品类目的总关联度;

将总关联度最大的商品类目,确定为包含该商品信息的节点对应的商品类目。

9. 一种商品信息与商品类目关联度的确定装置,其特征在于,包括:

节点确定模块,用于针对商品类目树中每个商品类目对应的节点,确定该节点的父节点;

控制模块,针对该节点包含的每个商品标题信息,进行以下步骤:对该商品标题信息分词,确定该商品标题信息对应的每个词组,其中所述词组包括至少一个分词,针对每个词组,根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值;

关联度确定模块,用于针对每个节点分别保存每个词组与该节点对应的商品类目的关联度值。

10. 一种基于权利要求 9 的关联度的确定装置确定商品类目信息的装置,其特征在于,该装置包括:

分词模块,用于接收获取的商品标题信息,对该商品标题信息进行分词,确定该商品标题信息对应的每个词组,其中该词组中包括至少一个分词;

查找模块,用于针对每个词组及与根节点连接的每个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度;

商品类目确定模块,用于根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点;将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

## 一种商品类目信息的确定方法及装置

### 技术领域

[0001] 本申请涉及数据处理技术领域,尤其涉及一种商品类目信息的确定方法及装置。

### 背景技术

[0002] 用户通过服务器上传商品的信息时,需要根据服务器提供的商品类目的信息,确定上传的该商品的信息所归属的类目,将该商品的信息挂靠在该类目下。但是由于服务器可以提供的商品类目的信息非常的庞杂,当用户主动确定商品的信息所归属的类目时,很可能错选类目,导致提供的信息不准确。

[0003] 另外,服务器采用开放搜索的方式,将其他服务器中保存的商品的信息纳入自身数据库中时,其他服务器中挂靠商品的信息的商品类目的划分类型可能与该服务器的商品类目的划分类型并不相同,在将其他服务器中保存的商品的信息纳入自身数据库时,如果商品类目的划分类型不同,将会导致其他服务器中保存的商品信息无法自动纳入到自身的数据库中,在这种情况下,只能采用人工识别的纳入方式,这大幅影响了商品信息纳入的效率及准确率。

[0004] 现有商品归属类目的确定方法包括:基于搜索的商品类目确定方法,以及长尾查询(query)类目确定方法。

[0005] 其中,基于搜索的商品类目确定方法中,当用户输入了商品的标题信息后,服务器根据用户输入的该商品的标题信息,通过搜索引擎在数据库中查找与该商品的标题信息完全匹配的商品的标题信息,用户根据查找到的该商品的标题信息匹配的商品挂靠的商品类目,确定输入的商品挂靠的类目。该方法中,只有查找到的标题信息完全匹配,才能确定商品所挂靠的商品类目,但是一般情况下,商品的标题信息包含的字符数很多,一般都在10~30个字符之间,因此,由于在大部分情况下都没办法查找到标题信息完全匹配的商品,并且逐一查找的过程非常耗时,这不但导致难以确定该商品挂靠的商品类目,而且效率低下。此外,即使查找到了完全匹配的标题信息,但当该标题信息对应的商品挂靠的商品类目存在错误时,也会导致确定的该商品挂靠的商品类目的出现差错。

[0006] 在长尾query类目确定方法中,当用户输入了查询关键词后,根据该查询关键词向用户返回查询结果,服务器根据用户输入的查询关键词,以及每个查询结果对应的商品类目,确定该查询关键词与每个商品类目的关联度,根据确定的关联度关系进行商品类目的确定。但是该方法在进行商品类目的确定时,采用的是一种扁平状的确定方法,即只确定了每个关键词与叶子节点对应的商品类目的关联度,在进行商品类目的确定过程中,需要逐一与每个叶子节点对应的商品类目进行匹配,而一般叶子节点对应的商品类目成千上万,这导致应用该方法确定商品类目的过程耗时较长,效率较低。

### 发明内容

[0007] 有鉴于此,本申请实施例提供一种商品类目信息的确定方法及装置,用以解决现有确定商品类目效率低,准确性差的问题。

- [0008] 本申请实施例提供的一种商品信息与商品类目关联度的确定方法,包括:
- [0009] 针对商品类目树中每个商品类目对应的节点,确定该节点的父节点;
- [0010] 针对该节点包含的每个商品标题信息,进行以下步骤:
- [0011] 对该商品标题信息分词,确定该商品标题信息对应的每个词组,其中所述词组包括至少一个分词;
- [0012] 针对每个词组,根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值;
- [0013] 针对该节点分别保存每个词组与该节点对应的商品类目的关联度值。
- [0014] 本申请实施例提供的一种基于上述关联度的确定方法确定商品类目信息的方法,包括:
- [0015] 接收商品标题信息;
- [0016] 对该商品标题信息进行分词,确定该商品标题信息对应的每个词组,其中每个词组中包括至少一个分词;
- [0017] 根据针对每个节点保存的每个词组与该节点对应的商品类目的关联度值,查找确定的每个词组与每个子节点的关联度;
- [0018] 根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点;
- [0019] 将确定的该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品信息挂靠到叶子节点对应的商品类目。
- [0020] 本申请实施例提供的一种商品信息与商品类目关联度的确定装置,包括:
- [0021] 节点确定模块,用于针对商品类目树中每个商品类目对应的节点,确定该节点的父节点;
- [0022] 控制模块,针对该节点包含的每个商品标题信息,进行以下步骤:对该商品标题信息分词,确定该商品标题信息对应的每个词组,其中所述词组包括至少一个分词,针对每个词组,根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值;
- [0023] 关联度确定模块,用于针对每个节点分别保存每个词组与该节点对应的商品类目的关联度值。
- [0024] 本申请实施例提供的一种基于上述关联度的确定装置确定商品类目信息的装置,包括:
- [0025] 分词模块,用于接收获取的商品标题信息,对该商品标题信息进行分词,确定该商品标题信息对应的每个词组,其中该词组中包括至少一个分词;
- [0026] 查找模块,用于针对每个词组及与根节点连接的每个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度;
- [0027] 商品类目确定模块,用于根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点;将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标

题信息对应的商品挂靠到叶子节点对应的商品类目。

[0028] 本申请实施例提供了一种商品类目信息的确定方法及装置,所述方法中,当获取到商品标题信息后,对该商品标题信息分词构成相应的词组,针对每个词组及与根节点连接的每个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度,并确定包含该商品信息的节点;将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。由于在本申请实施例中通过确定每个词组与每个商品类目的关联度,并且在确定的过程中从根节点开始依次查找包含商品信息的节点,由于与根节点连接的子节点的数量,远远小于商品类目树中包含的叶子节点的数量,因此可以大大缩短确定商品挂靠的叶子节点的时间,并且由于本申请实施例中,由与根节点连接的子节点依次确定包含该商品信息的子节点,从而可以保证商品挂靠的商品类目确定的准确性。

### 附图说明

[0029] 图 1A 为本申请实施例提供的一种商品类目树的结构示意图;

[0030] 图 1B 为本申请实施例提供的商品类目信息的确定过程;

[0031] 图 2A 为本申请实施例提供的商品信息与商品类目关联度的确定过程,涉及各个系统的组成结构示意图;

[0032] 图 2B 为本申请实施例提供的商品信息与商品类目关联度的确定过程;

[0033] 图 3 为本申请实施例提供的该 apache 模块具体的可以通过控制模块和确定模块实现结构示意图;

[0034] 图 4 为本申请实施例提供的商品类目信息的详细确定过程;

[0035] 图 5 为本申请实施例提供的一种商品信息与商品类目关联度的确定装置结构示意图;

[0036] 图 6 为本申请实施例提供的一种基于上述图 6 的关联度的确定装置确定商品类目信息的装置结构示意图。

### 具体实施方式

[0037] 本申请实施例为了提高确定商品所挂靠的商品类目的准确性及效率,提供了一种商品类目信息的确定方法、系统及装置,该方法通过确定每个词组与每个商品类目的关联度,并且在确定的过程中从根节点开始依次查找包含商品信息的节点,由于与根节点连接的子节点的数量,远远小于商品类目树中包含的叶子节点的数量,因此可以大大缩短确定商品挂靠的叶子节点的时间,并且由于本申请实施例中,由与根节点连接的子节点依次确定包含该商品信息的子节点,从而可以保证商品挂靠的商品类目确定的准确性。

[0038] 下面结合说明书附图,对本申请实施例进行详细说明。

[0039] 具体的,在本申请实施例中,数据库中保存有每个商品信息与其挂靠的商品类目信息的关系,该关系为用户根据其提供的商品信息与服务上提供的商品类目的对应关系而选择建立的。在该数据库中保存的该商品信息与其挂靠的商品类目信息的关系可以以以下格式存在:

[0040] 商品 ID- 商品标题信息 - 商品归属的商品类目 ID。

[0041] 其中,该商品归属的商品类目为商品类目树中的叶子节点对应的商品类目。在该商品类目树中,只存在该叶子节点的父节点等上层节点,不存在该叶子节点的子节点等下层节点,即该商品信息只能挂靠在叶子节点对应的商品类目下。

[0042] 另外,由于商品类目树中存在父节点与子节点之间的父子关系,当确定每个商品挂靠的商品类目时,通过与根节点连接的子节点依次确定对应的每个节点,最终挂靠到相应的叶子节点下的。因此在商品类目树中,挂靠了商品信息的叶子节点对应的父节点,一直到根节点都包含该商品的信息。

[0043] 图 1A 为本申请实施例提供的一种商品类目树的结构示意图,在该图 1A 中根节点的子节点包括“绿植园艺”“3C 数码”和“女装”等商品类目对应的节点,其中商品类目“绿植园艺”对应的节点的子节点包括“花卉盆栽”和“花种”商品类目的节点,另外,“花卉盆栽”商品类目对应的节点又包括子节点“球根花卉”以及“芳香花卉”对应的商品类目,并且“球根花卉”以及“芳香花卉”为叶子节点对应的商品类目,商品信息挂靠在叶子节点对应的商品类目下。当该商品信息挂靠在叶子类目下时,例如,挂靠在“球根花卉”商品类目对应的叶子节点下,此时,该叶子节点的父节点“花卉盆栽”包含该商品信息,并且“花卉盆栽”节点的父节点“绿植园艺”也包含该商品信息。这是因为,当某一商品信息挂靠在对应的商品类目对应的叶子节点下后,在数据库中除了保存该商品信息与其挂靠的商品类目信息,还保存有通过该商品类目树的节点到达该叶子节点的路径,因此,在数据库中保存的该商品信息与其挂靠的商品类目信息的关系可以以以下格式存在:商品 ID- 商品标题信息- 商品归属的商品类目的祖节点 ID- 商品归属的商品类目的父节点 ID- 商品归属的商品类目 ID。

[0044] 图 1B 为本申请实施例提供的商品类目信息的确定过程,该过程包括以下步骤:

[0045] S101:接收获取的商品标题信息,对该商品标题信息进行分词,确定该商品标题信息对应的每个词组,其中,每个词组中包括至少一个分词。

[0046] 其中,获取商品标题信息可以是,将接收到的用户输入的商品标题信息作为获取的商品标题信息,也可以是在将其他服务器的商品信息纳入自身数据库中时,将其他服务器中对该商品描述的商品标题信息,作为获取的商品标题信息。

[0047] 在本申请实施例中当获取了商品标题信息后,需对该商品标题信息进行分词,获取每个分词后,可以将至少一个分词组成一个词组。

[0048] S102:针对每个词组及与根节点连接的每个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度。

[0049] 在服务器中保存有每个词组与每个节点对应的商品类目的关联度值,即保存了该每个词组与每个商品类目的关联度值,当确定了每个词组后,可以根据保存的上述关联度值,确定该商品标题信息中每个词组与每个商品类目的关联度。

[0050] 具体的,当该词组中包含一个分词时,所述查找每个词组与每个子节点的关联度包括:针对每个词组,根据保存的每个词组与每个商品类目的关联度值,确定包含该词组的每个商品类目对应的子节点;针对确定的每个子节点,将保存的该词组与该子节点对应的商品类目的关联度值,作为查找到的该词组与该子节点的关联度。

[0051] S103:根据查找到的每个词组与每个子节点的关联度,确定包含该商品信息的节点。

[0052] 确定包含该商品信息的节点,包括:针对每个词组,根据查找到的该词组与每个子

节点的关联度,选择关联度较大的设定数量的子节点;针对选择的每个子节点对应的商品类目,根据该商品标题信息中每个词组与该商品类目的关联度,确定该商品标题信息与该商品类目的总关联度;将总关联度最大的商品类目,确定为包含该商品信息的节点对应的商品类目。

[0053] S104:将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

[0054] 本申请实施例提供的商品类目信息的确定方法,通过确定每个词组与每个商品类目的关联度,并且在确定的过程中从根节点开始依次查找包含商品信息的节点,由于与根节点连接的子节点的数量,远远小于商品类目树中包含的叶子节点的数量,因此可以大大缩短确定商品挂靠的叶子节点的时间,并且由于本申请实施例中,由与根节点连接的子节点依次确定包含该商品信息的子节点,从而可以保证商品挂靠的商品类目确定的准确性。

[0055] 在确定商品归属的商品类目之前,需要确定并保存每个词组与每个节点对应的商品类目的关联度值,该每个词组与商品类目的关联度值可以根据经验值设置,但是为了提高确定的商品类目信息的准确性,本申请实施例提供了一种商品信息与商品类目关联度的确定方法。

[0056] 图 2A 为本申请实施例提供的商品信息与商品类目关联度的确定及商品类目信息确定过程,由于在本申请实施例中当用户上传商品信息时,通过登录服务器将商品信息录入,服务器为用户提供其保存的商品类目树,用户通过由根节点不断的选择子孙节点,为其商品信息选择一个叶子节点对应的商品类目,此时在服务器中将该商品信息挂靠在该商品类目下,其中,该服务器可以为 WEB 服务器。当服务器将该商品信息挂靠在该商品类目下后,将该信息发送到数据库,由于涉及数据信息量巨大,因此可采用云计算平台实现,即由数据库周期性的将该信息导入到云计算平台,例如该云计算平台可以为 Hadoop。

[0057] 由于商品信息与其挂靠的商品类目信息的都保存到了云计算平台,因此云计算平台可以对自身保存的信息进行分析,确定每个词组与每个商品类目的关联度值。具体的云计算平台在确定每个词组与每个商品类目的关联度值时,可以采用分布式分析系统确定。

[0058] 云计算平台确定了每个词组与每个商品类目的关联度值后,将每个词组与每个商品类目的关联度的信息发送到服务器保存,以便服务器在确定商品类目信息时,可以根据该每个词组与每个商品类目的关联度值确定。

[0059] 图 2B 为本申请实施例提供的商品信息与商品类目关联度的确定过程,该过程包括以下步骤:

[0060] S201:针对商品类目树中每个商品类目对应的节点,确定该节点的父节点。

[0061] 具体的,在本申请实施例中,商品类目为每种商品所归属的种类,每个商品类目与商品类目树中的某个节点对应,最终挂靠商品信息的节点为叶子节点。并且商品类目树中由根节点开始,节点之间存在连接关系,对于根节点存在与其连接的子节点,对于叶子节点存在与其连接的父节点,对应中间节点即存在与其连接的父节点,也存在与其连接的子节点。

[0062] 当确定商品信息与商品类目的关联度时,为了保证确定的商品类目的准确性,需要针对每个节点确定相应词组与该节点的关联度值。

[0063] S202:针对该节点包含的每个商品标题信息,进行以下步骤:对该商品标题信息



分词,确定该商品标题信息对应的每个词组,其中所述词组包括至少一个分词,针对每个词组,根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值。

[0064] 具体的,确定该词组与该节点对应的商品类目的关联度值包括:根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,及该节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的第一权重值;根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该父节点对应的商品类目中其他商品类目的第二权重值;确定第一权重值与第二权重值的差值,根据该差值与第一权重值的商,确定该词组与该节点对应的商品类目的关联度值。

[0065] S203:针对每个节点分别保存每个词组与该节点对应的商品类目的关联度值。

[0066] 在本申请实施例中在确定词组与商品类目的关联度时,是针对该商品类目树中的每个节点对应的商品类目,分别确定该商品类目与相应词组的关联度。

[0067] 在确定每个词组与每个商品类目的关联度值时,由于每个节点对应的商品类目包含的商品信息是可知的,因此可以针对每个节点,确定该节点的父节点,针对该节点包含的每个商品的标题信息,确定该商品标题信息中的每个词组与该节点对应的商品类目的关联度值。并且在本申请实施例中可以针对每两个存在父子关系的节点,确定该子节点包含的每个商品的标题信息中每个词组与该子节点对应的商品类目的关联度值。

[0068] 仍以图 1A 为例,当确定每个词组与“花卉盆栽”节点的关联度时,根据该商品类目树可知该节点的父节点为“绿植园艺”节点,针对“花卉盆栽”节点包含的每个商品的标题信息,确定该商品标题信息中的每个词组与该“花卉盆栽”节点的关联度。当确定每个词组与“绿植园艺”节点关联度时,根据该商品类目树可知该节点的父节点为根节点,针对“绿植园艺”节点包含的每个商品的标题信息,确定该商品标题信息中的每个词组与该“绿植园艺”节点的关联度值。

[0069] 具体的在进行关联度值的计算时,可以以叶子节点到根节点的方向,依次确定每个节点与每个词组的关联度值,也可以以根节点到叶子节点的方向,依次确定每个节点与每个词组的关联度值。

[0070] 具体的在确定每个词组与每个商品类目的关联度值时,针对该商品类目对应的节点中包含的每个商品的标题信息,将该商品标题信息进行分词,将其中至少一个分词构成词组,确定该商品标题信息对应的每个词组,针对每个词组根据该词组与该节点的父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的关联度值。

[0071] 具体的确定该词组与该节点对应的商品类目的关联度值包括:

[0072] 根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,及该节点对应的商品类目包含的商品数量,确定该词组与该节点对应的商品类目的第一权重值;

[0073] 根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数,该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量,确定该词组与该父节点对应的商品类目中其他商品类目的第二权重值;

[0074] 确定第一权重值与第二权重值的差值,根据该差值与第一权重值的商,确定该词组与该节点对应的商品类目的关联度值。

[0075] 确定该词组与该节点对应的商品类目的第一权重值,包括:

[0076] 根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数,及该节点对应的商品类目包含商品数量的商值,确定第一权重值。

[0077] 下面以一个具体的例子进行说明。当商品 ID 为“1000”的商品信息挂靠在“球根花卉”商品类目对应的叶子节点下时,根据商品类目树,可知“球根花卉”商品类目对应的叶子节点,其父节点对应的商品类目为“花卉盆栽”。

[0078] 当商品 ID 为“1000”的商品信息的商品标题信息为“风信子水培紫色风信子种球可盆栽”,将该商品标题信息进行分词,针对分词“风信子”构成的词组,确定词组风信子与商品类目“球根花卉”的关联度值。

[0079] 例如与“花卉盆栽”商品类目关联的商品数量为 1623912,挂靠在“球根花卉”商品类目下的商品数量为 104286,并且词组“风信子”在“球根花卉”商品类目下的商品标题信息中出现的次数为 6588,在与“花卉盆栽”商品类目关联的的商品的商品标题信息中出现次数为 20683,则可知风信子相对于其所挂靠的“球根花卉”商品类目的第一权重值为  $6588/104286$ ,风信子与该“花卉盆栽”商品类目中其他商品类目的第二权重值为  $(20683-6588)/(1623912-104286)$ ,因此可知风信子与“球根花卉”商品类目的关联度值为  $[6588/104286-(20683-6588)/(1623912-104286)]/(6588/104286)$ ,即为 85.3%。

[0080] 由于该商品标题信息“风信子水培紫色风信子种球可盆栽”包含在“花卉盆栽”节点中,当确定该商品标题信息中每个词组与“花卉盆栽”节点的关联度值时,根据该“花卉盆栽”节点的父节点“绿植园艺”包含的每个商品的信息确定,其过程与上述过程相同,这里就不一一赘述。

[0081] 上述为确定词组与每个商品类目的关联度的过程,并且该过程是以该词组包括一个分词进行说明的,由于该分词中可以包括至少一个分词,当该词组中包含两个以上的分词时,将该两个以上的分词作为一个词组,根据上述确定方法,确定该词组与每个商品类目的关联度。该两个以上分词可以是该商品标题信息中的任意两个以上的分词,并且该两个以上分词的组成前后顺序不限于该商品标题信息中的该两个以上分词的前后顺序,例如当词组中包含两个分词时,该词组中的两个分词可以为“风信子”和“紫色”,也可以是“紫色”和“水培”,或者是“盆栽”和“水培”等,只要是任意两个分词构成的词组即可。这是因为对于商品标题信息中可能包括多个产品词对应的分词,根据上述方式确定每个分词与对应商品类目的关联度相差不大,为了进一步提高确定商品类目的准确性,需要结合包含至少两个分词的词组,与每个商品类目的关联度值,确定每个商品所挂靠的商品类目。

[0082] 例如对于 3C 行业中的数码产品和数码产品配件,经常出现商品标题信息多个分词与多个商品类目的关联度相当,仅依据单个分词与每个商品类目的关联度,很难通过商品标题信息判断一个商品究竟属于哪个商品类目。例如,当商品标题信息为“品胜正品三星

Galaxy S i9000 手机电池”，根据该商品标题信息，可以确定该商品的信息与商品类目“手机”和“手机电池”的关联度不相上下。为了解决这个问题，可以确定包含至少两个分词的词组，对于每个商品类目的关联度，例如确定词组“手机电池”与商品类目“手机电池”的关联度，通过该方法，可以确定该商品的信息与商品类目“手机电池”的关联度较大。

[0083] 这里为了简便，当该词组中包含一个分词时，可以将词组与每个商品类目的关联度值作为一阶表征向量，将至少两个分词组成的每个词组与每个商品类目的关联度值作为二阶表征向量。

[0084] 当确定并保存了每个词组与每个商品类目的关联度值后，获取了商品标题信息后，可以根据该商品标题信息，确定该商品标题信息对应的商品所归属的商品类目。

[0085] 由于将每个词组与每个商品类目的关联度保存在了服务器中，在确定每个商品归属的商品类目时，可以通过线上实时预测系统进行查找并确定每个商品所归属的商品类目，其中该线上实时预测系统可以位于服务器中，例如，为植入了 Apache 模块的 WEB 服务器。该 Apache 模块具体的可以通过控制模块和确定模块实现，如图 3 所示。

[0086] 在确定商品归属的商品类目时，可以是用户在将商品信息上传到服务器时，向服务器发送商品类目确定请求，当服务器接收到该商品类目确定请求后，提示用户输入该商品的商品标题信息，根据该商品标题信息，以及保存的每个词组与每个商品类目的关联度值，确定每个商品归属的商品类目。也可以是在服务器在将其他服务器中保存的商品信息纳入自身的数据库中时，识别并获取到每个商品的商品标题信息后，根据保存的每个词组与每个商品类目的关联度值，确定每个商品归属的商品类目。其具体的实现过程，可以根据需要灵活设定。

[0087] 在确定商品归属商品类目的过程中，当获取了商品标题信息后，为了提高确定的商品归属的商品类目的准确性，可以首先对该商品标题信息进行归一化处理，确定将该商品标题信息中的每个字符，对应的标准字符，例如当商品标题信息中存在英文的大写和小写时，服务器认为的标准字符为英文大写，则将该商品标题信息中的英文小写，调整为大写。之后将该商品标题信息进行分词，并根据保存的停用词信息表，过滤掉保存在该停用词信息表中的分词，针对过滤后得到的每个分词，确定该商品归属的商品类目。

[0088] 另外，由于在本申请实施例中可以确定每个词组与每个商品类目的关联度，该词组包括至少一个分词，因此在服务器在保存每个词组与每个商品类目的关联度值时，可以针对一个分词，两个或多个分词分别保存每个词组与每个商品类目的关联度值，即在服务器中保存有关联度关系表，其中一个表中记录的为每个分词与每个商品类目的关联度值，另一个表中记录的为由至少两个分词构成的词组与每个商品类目的关联度值。

[0089] 当确定每个商品归属的商品类目时，由于已经确定了每个商品标题信息中的每个分词，将至少一个分词组成一个词组，可以首先根据与根节点连接的该根节点的每个子节点，根据保存的该每个子节点与每个词组的关联度值，确定每个词组与该每个子节点的关联度，根据该商品标题信息中每个词组与该每个子节点的关联度，确定包含该商品标题信息的子节点，之后将该子节点作为父节点，确定该父节点的子节点，依据上述步骤，确定该商品标题信息中每个词组与该每个子节点的关联度，确定包含该商品标题信息的子节点，直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

[0090] 并且在确定该商品标题信息对应的商品挂靠的叶子节点对应的商品类目时，可以

首先针对每个分词,分别查找每个分词与每个商品类目的关联度。为了提高确定的商品类目的准确性,只有在某一分词与某一商品类目的关联度大于设定的关联度阈值时,才确定该分词与其他分词构成的词组,查找该词组与每个商品类目的关联度。

[0091] 图 4 为本申请实施例提供的商品归属商品类目的详细确定过程,该过程包括以下步骤:

[0092] S401:服务器接收用户所在的客户端发送的商品类目确定请求;

[0093] S402:服务器根据接收到的商品类目确定请求,提示并获取用户输入的商品标题信息;

[0094] S403:对该商品标题信息进行归一化、分词及过滤处理,得到该商品标题信息对应的每个分词;

[0095] S404:提取该商品标题信息对应的分词中一个,将提取的该分词作为当前关联度确定的分词,针对该提取的分词,以及与根节点连接的每个子节点,根据保存的每个词组与该每个子节点对应的商品类目的关联度值,查找该词组与每个子节点的关联度;

[0096] S405:判断查找到的该分词与每个子节点对应的商品类目的关联度中,是否存在超过设定的关联度阈值的关联度,当存在时,进行步骤 S407,否则,进行步骤 S406;

[0097] S406:判断该分词是否为该商品标题信息中的最后一个分词,当判断结果为是时,进行步骤 S408,否则,转步骤 S404;

[0098] S407:将该分词与该分词之后的至少一个分词构成词组,针对该词组,以及与根节点连接的每个子节点,根据保存的每个词组与该每个子节点对应的商品类目的关联度值,查找该词组与每个子节点的关联度;

[0099] S408:针对查找的每个词组与每个子节点的关联度,选择关联度较大的设定数量的子节点;

[0100] S409:针对选择的每个子节点对应的商品类目,根据该商品标题信息中每个分词与该商品类目的关联度,确定该商品标题信息与该商品类目的第一总关联度,根据每个词组与该商品类目的关联度,确定该商品标题信息与该商品类目的第二总关联度;

[0101] S410:根据该商品标题信息与每个商品类目的第一总关联度和第二总关联度,确定该商品标题信息与每个商品类目的总关联度,将总关联度最大的商品类目,确定为包含该商品信息的节点对应的商品类目;

[0102] S411:将包含该商品信息的节点作为父节点,确定该父节点的子节点,返回步骤 S404,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

[0103] 具体的例如,当该商品的标题信息包括分词 A、B、C 时,该设定数量的商品类目为 3,下表为该商品标题信息中包含的每个分词与每个商品类目的关联度值:

分词	商品 类目	R1	R2	R3	R4	R5
	A		50%	83%	62%	
B		40%		20%		37%
C		42%	57%		32%	
AB		20%	40%			30%
AC		10%	30%	40%		

[0104] 根节点包含的子节点对应的商品类目分别为 R1、R2、R3，并且确定分词 A 与商品类目 R1、R2、R3 的关联度值分别为 50%、83% 和 62%，分词 B 与商品类目 R1、R3、R5 的关联度值分别为 40%、20% 和 37%，分词 C 与商品类目 R1、R2、R4 的关联度值分别为 42%、57% 和 32%，该设定的关联度阈值为 70%，确定分词 A 与分词 B 构成的词组与商品类目 R1、R2、R5 的关联度值分别为 20%、40% 和 30%，分词 A 与分词 C 构成的词组与商品类目 R1、R2、R3 的关联度值分别为 10%、30% 和 40%。则可以确定该商品标题信息与商品类目 R1 的第一总关联度值为 132%，该商品标题信息与商品类目 R2 的第一总关联度值为 140%，该商品标题信息与商品类目 R3 的第一总关联度值为 82%，该商品标题信息与商品类目 R4 的第一总关联度值为 32%，该商品标题信息与商品类目 R5 的第一总关联度值为 37%。

[0105] 另外，也可以确定该商品标题信息与商品类目 R1 的第二总关联度值为 30%，该商品标题信息与商品类目 R2 的第二总关联度值为 70%，商品标题信息与商品类目 R3 的第二总关联度值为 40%，商品标题信息与商品类目 R5 的第二总关联度值为 30%。

[0106] 因此该商品标题信息与商品类目 R1 的总关联度值为 162%，该商品标题信息与商品类目 R2 的总关联度值为 210%，该商品标题信息与商品类目 R3 的总关联度值为 122%，该商品标题信息与商品类目 R4 的总关联度值为 32%，该商品标题信息与商品类目 R5 的总关联度值为 67%，因此可知包含该商品信息的节点为商品类目 R2 对应的节点。之后以该节点为父节点，确定该父节点的子节点，在该子节点中采用上述方法，确定包含该商品信息的节点，之后再以该节点为父节点，确定该父节点的子节点依次采用上述方法，直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

[0107] 图 5 为本申请实施例提供的一种商品信息与商品类目关联度的确定装置结构示意图，该装置包括：

[0108] 节点确定模块 51，用于针对商品类目树中每个商品类目对应的节点，确定该节点的父节点；

[0109] 控制模块 52，针对该节点包含的每个商品标题信息，进行以下步骤：对该商品标题信息分词，确定该商品标题信息对应的每个词组，其中所述词组包括至少一个分词，针对每个词组，根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数，该词组在该节点对应的商品类目包含的商品标题信息中出现的次数，该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量，确定该词组与该节点对应的

商品类目的关联度值；

[0111] 关联度确定模块 53, 用于针对每个节点分别保存每个词组与该节点对应的商品类目的关联度值。

[0112] 所述控制模块 52, 具体用于根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数, 及该节点对应的商品类目包含的商品数量, 确定该词组与该节点对应的商品类目的第一权重值; 根据该词组在该父节点对应的商品类目包含的商品标题信息中出现次数, 该词组在该节点对应的商品类目包含的商品标题信息中出现的次数, 该节点对应的商品类目包含的商品数量及与该父节点对应的商品类目包含的商品数量, 确定该词组与该父节点对应的商品类目中其他商品类目的第二权重值; 确定第一权重值与第二权重值的差值, 根据该差值与第一权重值的商, 确定该词组与该节点对应的商品类目的关联度值。

[0113] 所述控制模块 52, 具体用于根据该词组在该节点对应的商品类目包含的商品标题信息中出现的次数, 及该节点对应的商品类目包含商品数量的商值, 确定第一权重值。

[0114] 所述控制模块 52, 具体用于确定该词组在该父节点对应的商品类目包含的商品标题信息中出现次数, 及该词组在该节点对应的商品类目包含的商品标题信息中出现的次数的第一差值; 确定该父节点对应的商品类目包含的商品数量, 及该节点对应的商品类目包含的商品数量的第二差值; 根据第一差值与第二差值的商值, 确定该第二权重值。

[0115] 图 6 为本申请实施例提供的一种基于上述图 6 的关联度的确定装置确定商品类目信息的装置结构示意图, 该装置包括:

[0116] 分词模块 61, 用于接收获取的商品标题信息, 对该商品标题信息进行分词, 确定该商品标题信息对应的每个词组, 其中该词组中包括至少一个分词;

[0117] 查找模块 62, 用于针对每个词组及与根节点连接的每个子节点, 根据保存的每个词组与每个节点对应的商品类目的关联度值, 查找每个词组与每个子节点的关联度;

[0118] 商品类目确定模块 63, 用于根据查找到的每个词组与每个子节点的关联度, 确定包含该商品信息的节点; 将该子节点作为父节点, 确定该父节点的子节点, 直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。

[0119] 所述查找模块 62, 具体用于当该词组中包含一个分词时, 针对每个词组, 根据保存的每个词组与每个商品类目的关联度值, 确定包含该词组的每个商品类目对应的子节点; 针对确定的每个子节点, 将保存的该词组与该子节点对应的商品类目的关联度, 作为查找到的该词组与该子节点的关联度。

[0120] 所述查找模块 62, 还用于判断查找到的该词组与该子节点的关联度是否大于设定的关联度阈值; 当判断结果为是时, 将所述分词与该商品标题信息中的其他分词组成词组, 查找该词组与每个商品类目的关联度。

[0121] 所述商品类目确定模块 63, 具体用于针对每个词组, 根据查找到的该词组与每个子节点的关联度, 选择关联度较大的设定数量的子节点; 针对选择的每个子节点对应的商品类目, 根据该商品标题信息中每个词组与该商品类目的关联度, 确定该商品标题信息与该商品类目的总关联度; 将总关联度最大的商品类目, 确定为包含该商品信息的节点对应的商品类目。

[0122] 本申请实施例提供了一种商品类目信息的确定方法及装置, 方法中当获取到商品标题信息后, 对该商品标题信息分词构成相应的词组, 针对每个词组及与根节点连接的每

个子节点,根据保存的每个词组与每个节点对应的商品类目的关联度值,查找每个词组与每个子节点的关联度,并确定包含该商品信息的节点;将该子节点作为父节点,确定该父节点的子节点,直到将所述商品标题信息对应的商品挂靠到叶子节点对应的商品类目。由于在本申请实施例中通过确定每个词组与每个商品类目的关联度,并且在确定的过程中从根节点开始依次查找包含商品信息的节点,由于与根节点连接的子节点的数量,远远小于商品类目树中包含的叶子节点的数量,因此可以大大缩短确定商品挂靠的叶子节点的时间,并且由于本申请实施例中,由与根节点连接的子节点依次确定包含该商品信息的子节点,从而可以保证商品挂靠的商品类目确定的准确性。

[0123] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

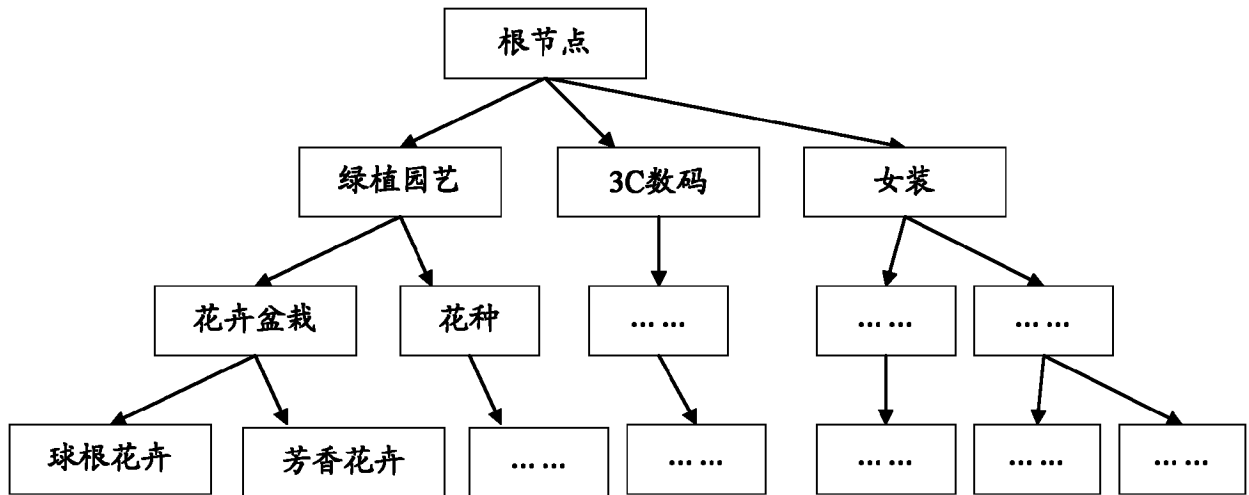


图 1A

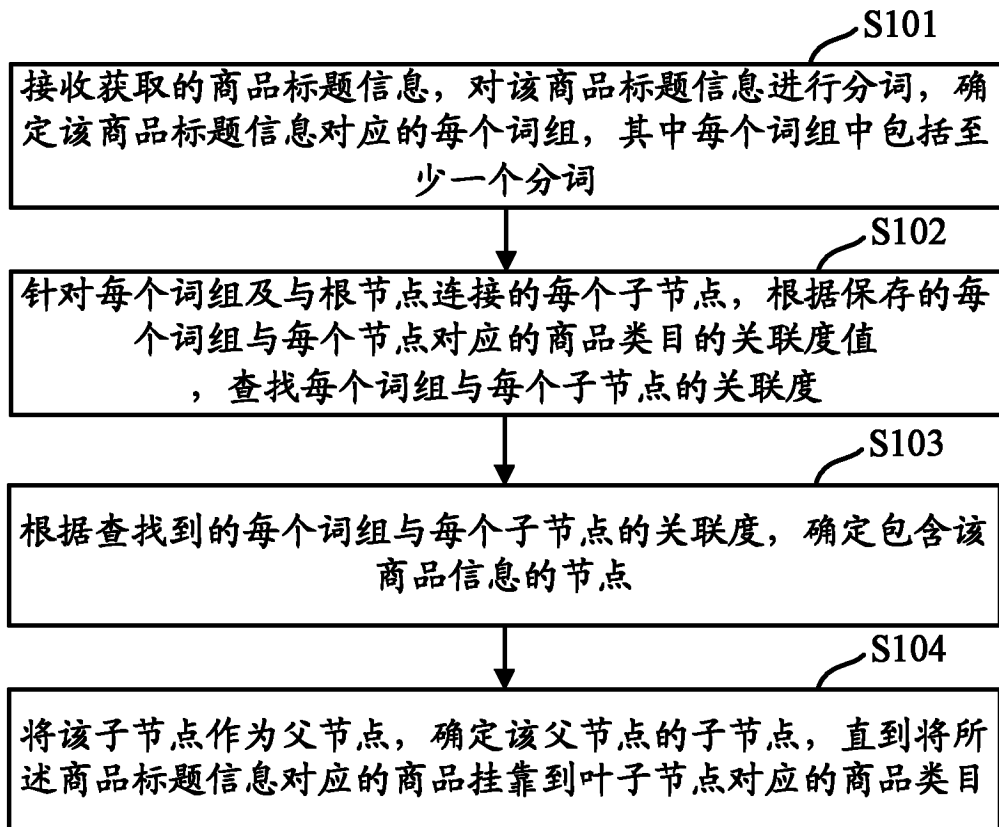


图 1B



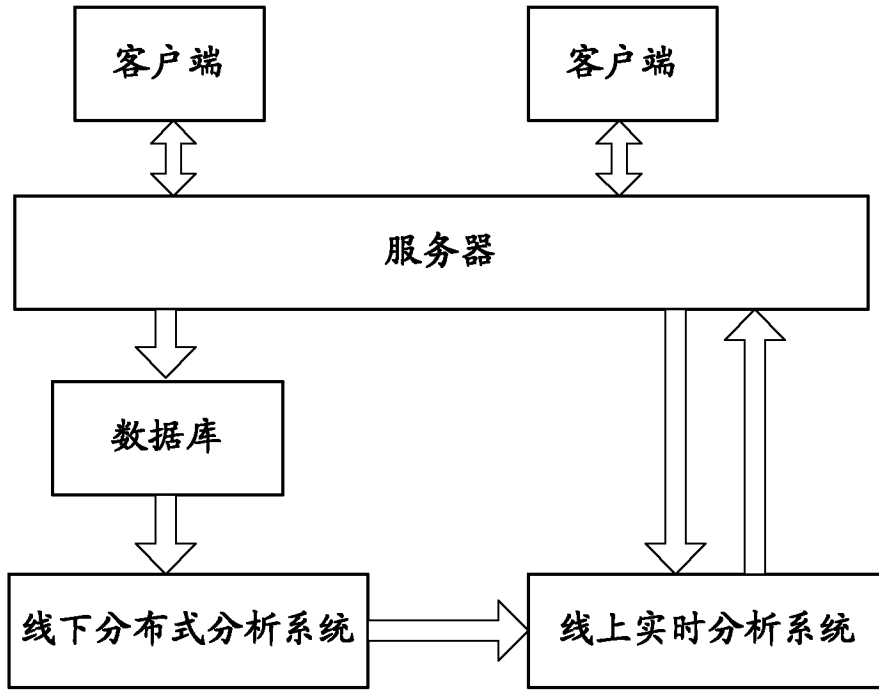


图 2A

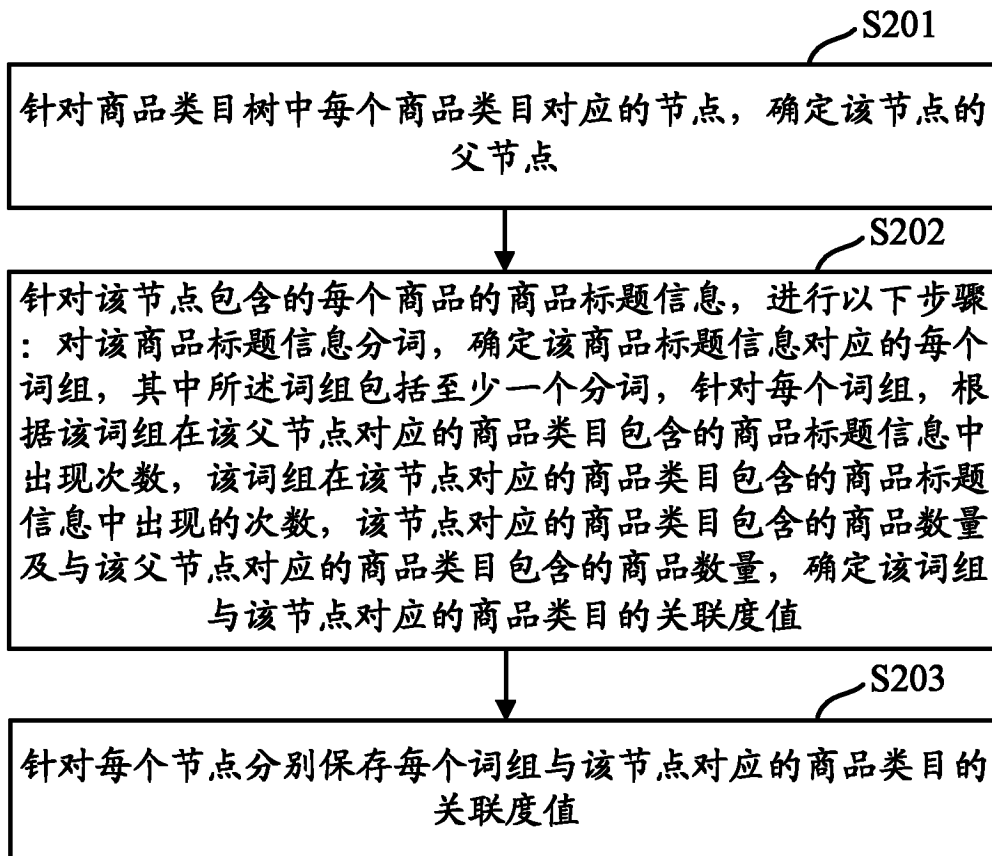


图 2B

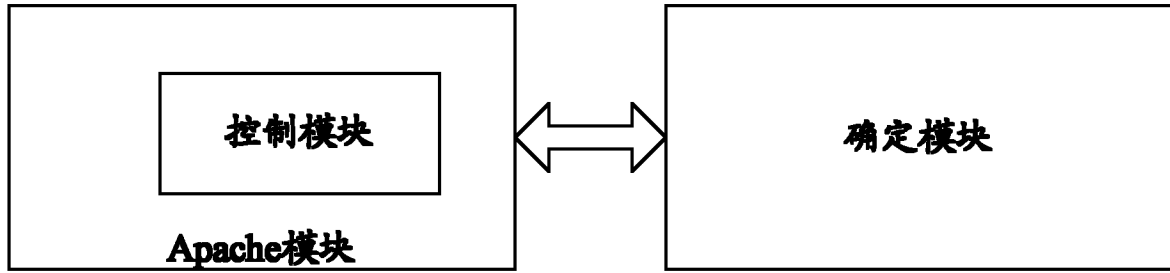


图 3

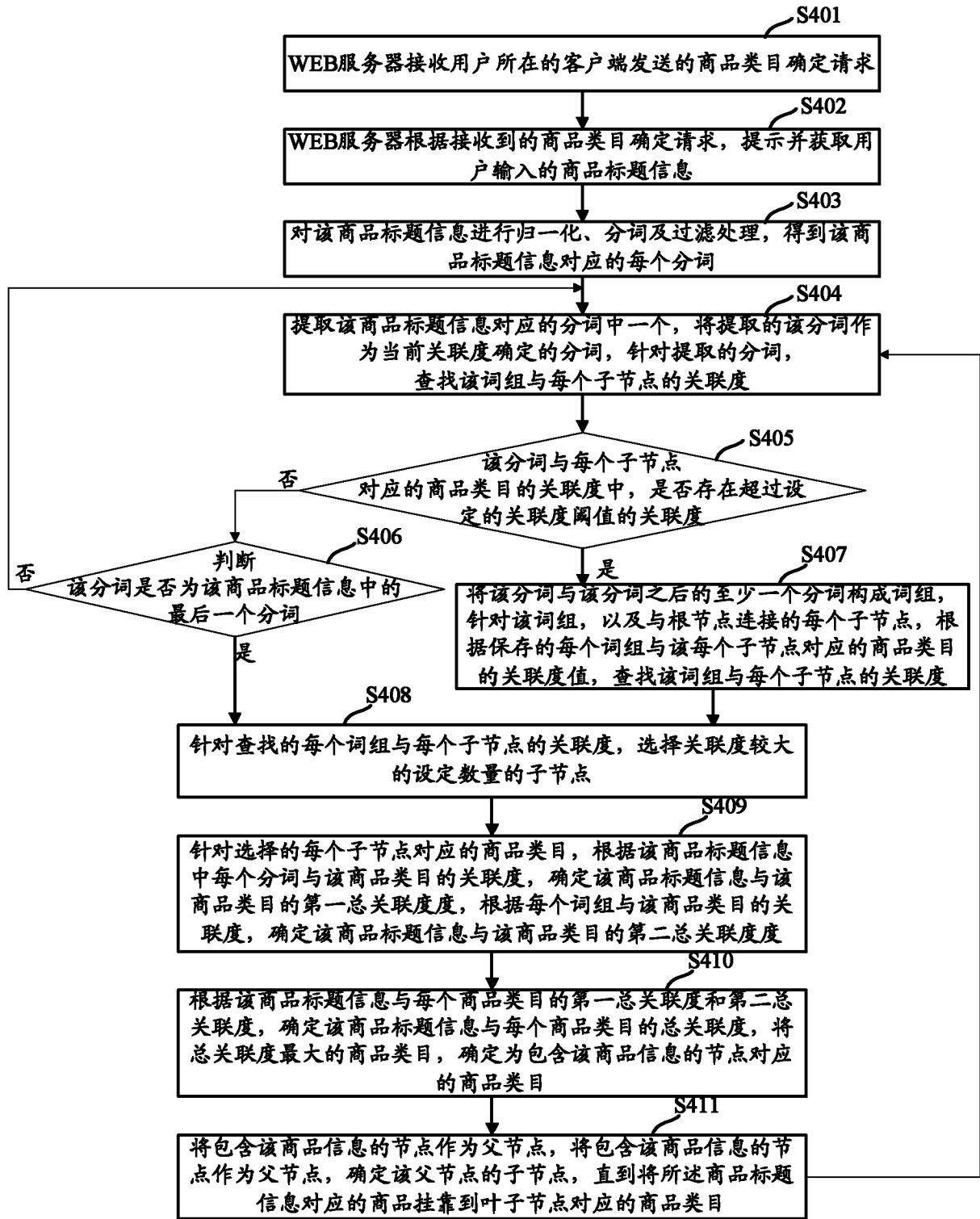


图 4

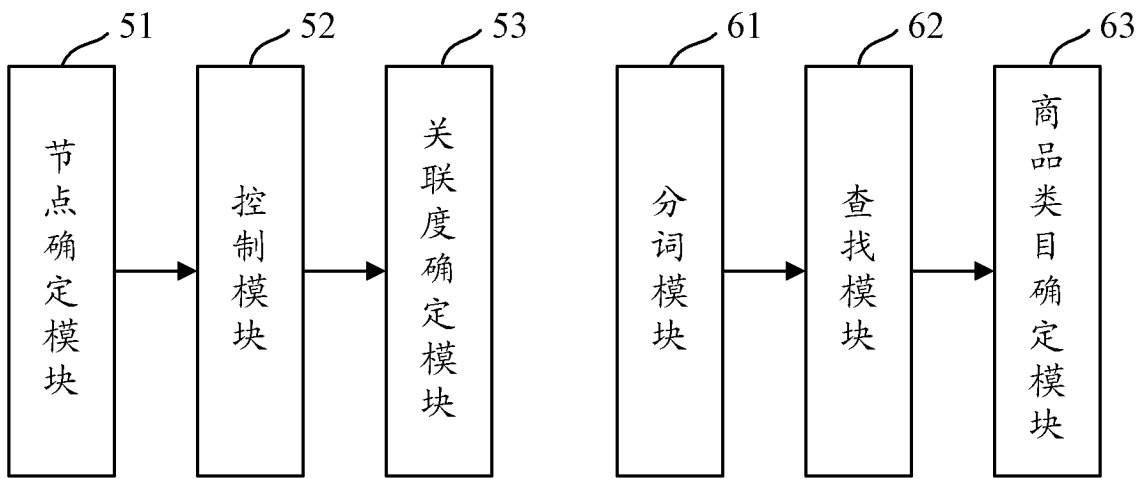


图 5

图 6