



(12) 发明专利申请

(10) 申请公布号 CN 114140077 A

(43) 申请公布日 2022.03.04

(21) 申请号 202111441163.X

G06F 40/211 (2020.01)

(22) 申请日 2021.11.30

G06F 40/237 (2020.01)

G06F 40/289 (2020.01)

(71) 申请人 宁波帮企一把企业服务平台有限公司

地址 315046 浙江省宁波市鄞州区朝晖路
17号上海银行大厦9楼(9-1)-(9-4)

(72) 发明人 孙小健 谭维定 王红星 傅倩倩

(74) 专利代理机构 上海天翔知识产权代理有限公司 31224

代理人 陈骏键

(51) Int. Cl.

G06Q 10/10 (2012.01)

G06Q 50/26 (2012.01)

G06F 40/216 (2020.01)

G06F 16/335 (2019.01)

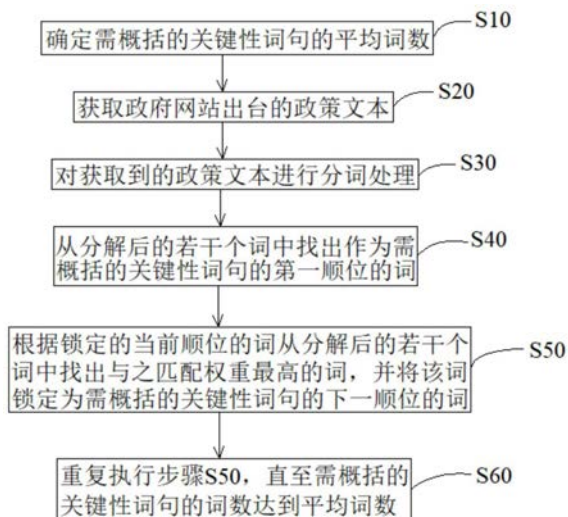
权利要求书3页 说明书10页 附图10页

(54) 发明名称

一种政府政策解构方法、装置、计算机设备和存储介质

(57) 摘要

本发明公开的一种政府政策解构方法,包括以下步骤:1.确定需概括的关键性词句的平均词数;2.获取政府网站出台的政策文本;3.对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;4.从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;5.根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;6.重复执行步骤5,直至需概括的关键性词句的词数达到平均词数。还公开了实现上述方法的装置、计算机设备和存储介质。本发明的概括准确率高,提高了工作效率。



1. 一种政府政策解构方法,其特征在于,包括以下步骤:

步骤S10,确定需概括的关键性词句的平均词数;

步骤S20,获取政府网站出台的政策文本;

步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;

步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;

步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;

步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。

2. 如权利要求1所述的政府政策解构方法,其特征在于,在步骤S10中,所述确定需概括的关键性词句的平均词数,包括以下步骤:

步骤S11,获取历史数据库中所有的关键性词句;

步骤S12,对每一关键性词句的词数进行分词处理,使得每一关键性词句被分解成若干个词;

步骤S13,对分词处理后的每一关键性词句的词数进行统计,并对统计后的数据进行平均计算处理,计算所得到的数值为需概括的关键性词句的平均词数。

3. 如权利要求1所述的政府政策解构方法,其特征在于,在步骤S30中,所述对获取到的政策文本进行分词处理,包括以下步骤:

步骤S31,对获取到的政策文本进行分段处理,使得所述政策文本被分解成若干个段落;

步骤S32,对分解后的每一个段落进行分句处理,使得每一个段落被分解成若干个句子;

步骤S33,对分解后的每一个句子进行分词处理,使得每一个句子被分解成若干个词。

4. 如权利要求1所述的政府政策解构方法,其特征在于,在步骤S40中,所述从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词,包括以下步骤:

步骤S41,从词顺序库内找出第一顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入步骤S42,若匹配失败,则进入步骤S43;

步骤S42,将匹配到的词锁定为需概括的关键性词句的第一顺位的词;

步骤S43,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则进入步骤S44;

步骤S44,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则返回步骤S43。

5. 如权利要求1所述的政府政策解构方法,其特征在于,在步骤S50中,所述根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,包括以下步骤,

步骤S51,根据锁定的当前顺位的词从神经元库内找出与之匹配权重最高的词,并将该

词与分解后的若干个词进行逐一匹配,若匹配成功,则进入步骤S52,若匹配失败,则进入步骤S53;

步骤S52,将匹配到的词锁定为需概括的关键性词句的下一顺位的词;

步骤S53,降低匹配权重,并根据锁定的当前顺位的词从神经元库内找出与之降低后的匹配权重对应的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S54;

步骤S54,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入返回步骤S52,若匹配失败,则进入步骤S55;

步骤S55,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S56;

步骤S56,从词顺序库内找出当前顺位的下一顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则返回步骤S54。

6. 如权利要求1至5中任一项所述的政府政策解构方法,其特征在于,还包括步骤S70,对概括形成的关键性词句进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,返回步骤S10,对关键性词句进行重新构建。

7. 如权利要求6所述的政府政策解构方法,其特征在于,还包括步骤S80,对语法判断处理后的关键性词句进行人工校验,若人工校验成功,则将概括形成的关键性词句推送至目标者,若人工校验失败,则返回步骤S10,对关键性词句进行重新构建。

8. 如权利要求7所述的政府政策解构方法,其特征在于,还包括步骤S90,对人工校验成功的关键性词句进行自我学习处理。

9. 如权利要求8所述的政府政策解构方法,其特征在于,在步骤S90中,对人工校验成功的关键性词句进行自我学习处理,包括以下步骤:

步骤S91,对人工校验成功的关键性词句进行分词处理,使得关键性词句被分解为若干个词;

步骤S92,建立每相邻两个词之间的匹配权重,并将建立后的匹配权重保存至神经元库内;

步骤S93,对分解后的每个词进行词性识别,并将识别结果保存至词性库内;

步骤S94,从互联网上获取分解后的每个词的近似词,并将获取到的这些近义词保存至近义词库内;

步骤S95,对分解后的每个词进行优先级排序处理,并将优先级排序结果保存至词顺序库内。

10. 一种政府政策关键性信息构建装置,其特征在于,包括:

平均词数计算模块,所述平均词数计算模块用于确定需概括的关键性词句的平均词数;

政策文本获取模块,所述政策文本获取模块用于获取政府网站出台的政策文本;

分词处理模块,所述分词处理模块用于对获取到的政策文本进行分词处理,以使得所

述政策文本被分解成若干个词；

第一词句匹配模块,所述第一词句匹配模块用于从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;以及

第二词句匹配模块,所述第二词句匹配模块用于根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词,如此重复,直至需概括的关键性词句的词数达到平均词数。

11. 如权利要求10所述的政府政策关键性信息构建装置,其特征在于,还包括语法判断处理模块,所述语法判断处理模块用于对概括形成的关键性词句进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,对关键性词句进行重新构建。

12. 如权利要求10所述的政府政策关键性信息构建装置,其特征在于,还包括自我学习模块,所述自我学习模块用于对概括形成的关键性词句进行自我学习处理。

13. 一种用于实现上述政府政策解构方法的计算机设备,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如权利要求1至9中任一项所述的政府政策解构方法中的步骤。

14. 一种用于实现上述政府政策解构方法的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如权利要求1至9中任一项所述的政府政策解构方法的步骤。

一种政府政策解构方法、装置、计算机设备和存储介质

技术领域

[0001] 本发明涉及电子政务技术领域,尤其涉及一种政府政策解构方法、装置、计算机设备和存储介质。

背景技术

[0002] 政策为以权威形式标准化地规定在一定的历史时期内,应该达到的奋斗目标、遵循的行动原则、完成的明确任务、实行的工作方式、采取的一般步骤和具体措施。政策具有以下特点:1.时效性。政策是在一定时间内的历史条件和国情条件下,推行的现实政策。2.表述性。就表现形态而言,政策不是物质实体,而是外化为符号表达的观念和信息,它由有权机关用语言和文字等表达手段进行表述。

[0003] 随着网络技术的快速发展,催生出一种电子政务技术。电子政务,其为运用计算机、网络和通信等现代信息技术手段,实现政府组织结构和 workflows 的优化重组,超越时间、空间和部门分隔的限制,建成一个精简、高效、廉洁、公平的政府运作模式,以便全方位地向社会提供优质、规范、透明、符合国际水准的管理与服务。

[0004] 政府作为国家管理部门,其本身上网开展电子政务,有助于政府管理的现代化,实现政府办公电子化、自动化、网络化。通过互联网这种快捷、廉价的通信手段,政府可以让公众迅速了解政府机构的组成、职能和办事章程,以及各项政策法规,增加办事执法的透明度,并自觉接受公众的监督。

[0005] 在电子政务中,政府机关的各种数据、文件、档案、社会经济数据都以数字形式存储于网络服务器中,可通过计算机检索机制快速查询、即用即调。

[0006] 虽然电子政务的推广使用,使得人们能够及时、快速的获取到政府所发布的政策,但是企业或个人若想办理相关政策只有在知道有这项政策的情况下才会去了解 and 办理相关的政策,然而从中央到地方政策种类和内容都非常的多,企业或个人不可能都对每个政策都了解。同时,大多数政策的办理都是有时效性的,尤其是一些资金性政策,一旦错过办理时限,则不能享受到政策所给予的资金奖励或者优惠。因此,当政策正式发布后,应当及时有效地将政策推送给可能符合条件的企业或个人。通常而言,政策文件中的内容较多,如果直接将政策文件中的所有内容推送给目标企业或个人,显然对目标企业或个人而言是难以把握重点信息的。因此,需要将政策文件中的目标企业或个人感兴趣的关键性信息提取后,再将这些关键性信息推送给目标企业或个人,这样才能更好地激发目标企业或个人办理项目的主动性,同时也有利于政策的实施。现有的做法大多是通过人工方式提取政策文件中的关键性信息,但由于各类政策文件出台较多且每个政策文件的内容较多,仅依靠人工方式提取,效率低下,耗时长,同时也增加了人力成本。

[0007] 为此,本申请人经过有益的探索和研究,找到了解决上述问题的方法,下面将要介绍的技术方案便是在这种背景下产生的。

发明内容

[0008] 本发明所要解决的技术问题之一在于：针对现有技术的不足而提供一种提高效率、节省时间、降低人力成本的政府政策解构方法。

[0009] 本发明所要解决的技术问题之二在于：提供一种实现上述政府政策解构方法的政府政策关键性信息构建装置。

[0010] 本发明所要解决的技术问题之三在于：提供一种实现上述政府政策解构方法的计算机设备。

[0011] 本发明所要解决的技术问题之四在于：提供一种实现上述政府政策解构方法的计算机可读存储介质。

[0012] 作为本发明第一方面的一种政府政策解构方法，包括以下步骤：

[0013] 步骤S10，确定需概括的关键性词句的平均词数；

[0014] 步骤S20，获取政府网站出台的政策文本；

[0015] 步骤S30，对获取到的政策文本进行分词处理，以使得所述政策文本被分解成若干个词；

[0016] 步骤S40，从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词；

[0017] 步骤S50，根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词，并将该词锁定为需概括的关键性词句的下一顺位的词，此时该词成为需概括的关键性词句的当前顺位的词；

[0018] 步骤S60，重复执行步骤S50，直至需概括的关键性词句的词数达到平均词数。

[0019] 在本发明的一个优选实施例中，在步骤S10中，所述确定需概括的关键性词句的平均词数，包括以下步骤：

[0020] 步骤S11，获取历史数据库中所有的关键性词句；

[0021] 步骤S12，对每一关键性词句的词数进行分词处理，使得每一关键性词句被分解成若干个词；

[0022] 步骤S13，对分词处理后的每一关键性词句的词数进行统计，并对统计后的数据进行平均计算处理，计算所得到的数值为需概括的关键性词句的平均词数。

[0023] 在本发明的一个优选实施例中，在步骤S30中，所述对获取到的政策文本进行分词处理，包括以下步骤：

[0024] 步骤S31，对获取到的政策文本进行分段处理，使得所述政策文本被分解成若干个段落；

[0025] 步骤S32，对分解后的每一个段落进行分句处理，使得每一个段落被分解成若干个句子；

[0026] 步骤S33，对分解后的每一个句子进行分词处理，使得每一个句子被分解成若干个词。

[0027] 在本发明的一个优选实施例中，在步骤S40中，所述从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词，包括以下步骤：

[0028] 步骤S41，从词顺序库内找出第一顺位的词，并将该词与分解后的若干个词进行逐一匹配，若匹配成功，则进入步骤S42，若匹配失败，则进入步骤 S43；

[0029] 步骤S42，将匹配到的词锁定为需概括的关键性词句的第一顺位的词；

[0030] 步骤S43,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则进入步骤S44;

[0031] 步骤S44,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则返回步骤S43。

[0032] 在本发明的一个优选实施例中,在步骤S50中,所述根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,包括以下步骤,

[0033] 步骤S51,根据锁定的当前顺位的词从神经元库内找出与之匹配权重最高的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入步骤S52,若匹配失败,则进入步骤S53;

[0034] 步骤S52,将匹配到的词锁定为需概括的关键性词句的下一顺位的词;

[0035] 步骤S53,降低匹配权重,并根据锁定的当前顺位的词从神经元库内找出与之降低后的匹配权重对应的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S54;

[0036] 步骤S54,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入返回步骤S52,若匹配失败,则进入步骤S55;

[0037] 步骤S55,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S56;

[0038] 步骤S56,从词顺序库内找出当前顺位的下一顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则返回步骤S54。

[0039] 在本发明的一个优选实施例中,还包括步骤S70,对概括形成的关键性词句进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,返回步骤S10,对关键性词句进行重新构建。

[0040] 在本发明的一个优选实施例中,还包括步骤S80,对语法判断处理后的关键性词句进行人工校验,若人工校验成功,则将概括形成的关键性词句推送至目标者,若人工校验失败,则返回步骤S10,对关键性词句进行重新构建。

[0041] 在本发明的一个优选实施例中,还包括步骤S90,对人工校验成功的关键性词句进行自我学习处理。

[0042] 在本发明的一个优选实施例中,在步骤S90中,对人工校验成功的关键性词句进行自我学习处理,包括以下步骤:

[0043] 步骤S91,对人工校验成功的关键性词句进行分词处理,使得关键性词句被分解为若干个词;

[0044] 步骤S92,建立每相邻两个词之间的匹配权重,并将建立后的匹配权重保存至神经元库内;

- [0045] 步骤S93,对分解后的每个词进行词性识别,并将识别结果保存至词性库内;
- [0046] 步骤S94,从互联网上获取分解后的每个词的近似词,并将获取到的这些近义词保存至近义词库内;
- [0047] 步骤S95,对分解后的每个词进行优先级排序处理,并将优先级排序结果保存至词顺序库内。
- [0048] 作为本发明第二方面的一种政府政策关键性信息构建装置,包括:
- [0049] 平均词数计算模块,所述平均词数计算模块用于确定需概括的关键性词句的平均词数;
- [0050] 政策文本获取模块,所述政策文本获取模块用于获取政府网站出台的政策文本;
- [0051] 分词处理模块,所述分词处理模块用于对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;
- [0052] 第一词句匹配模块,所述第一词句匹配模块用于从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;以及
- [0053] 第二词句匹配模块,所述第二词句匹配模块用于根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词,如此重复,直至需概括的关键性词句的词数达到平均词数。
- [0054] 在本发明的一个优选实施例中,还包括语法判断处理模块,所述语法判断处理模块用于对概括形成的关键性词句进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,对关键性词句进行重新构建。
- [0055] 在本发明的一个优选实施例中,还包括自我学习模块,所述自我学习模块用于对概括形成的关键性词句进行自我学习处理。
- [0056] 作为本发明第三方面的一种用于实现上述政府政策解构方法的计算机设备,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:
- [0057] 步骤S10,确定需概括的关键性词句的平均词数;
- [0058] 步骤S20,获取政府网站出台的政策文本;
- [0059] 步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;
- [0060] 步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;
- [0061] 步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;
- [0062] 步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。
- [0063] 作为本发明第四方面的一种用于实现上述政府政策解构方法的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:
- [0064] 步骤S10,确定需概括的关键性词句的平均词数;
- [0065] 步骤S20,获取政府网站出台的政策文本;

[0066] 步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;

[0067] 步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;

[0068] 步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;

[0069] 步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。

[0070] 由于采用了如上技术方案,本发明的有益效果在于:本发明对政策文本进行分词处理,再从分解后的若干个词中逐一找出需概括的关键性词句中的每一个词,使得找出来的这些词构建概括后的关键性词句,概括准确率高,有效地提高了工作效率,节省了工作时间,降低了人力成本。

附图说明

[0071] 为了更清楚地说明本发明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的前提下,还可以根据这些附图获得其他的附图。

[0072] 图1是本发明的政府政策解构方法的一种实施例的应用场景图。

[0073] 图2是本发明的政府政策解构方法的一种实施例的流程图。

[0074] 图3是本发明的确定平均词数的流程图。

[0075] 图4是本发明的分词处理的流程图。

[0076] 图5是本发明的找出作为需概括的关键性词句的第一顺位的词的流程图。

[0077] 图6是本发明的找出锁定为需概括的关键性词句的下一顺位的词的流程图。

[0078] 图7是本发明的政府政策解构方法的一种具体应用实施例的流程示意图。

[0079] 图8是本发明的政府政策解构方法的另一种实施例的流程图。

[0080] 图9是本发明的政府政策解构方法的又一种实施例的流程图。

[0081] 图10是本发明的政府政策解构方法的又一种实施例的流程图。

[0082] 图11是本发明的政府政策解构方法中的自我学习处理的流程图。

[0083] 图12是本发明的政府政策关键性信息构建装置的一种实施例的结构示意图。

[0084] 图13是本发明的计算机设备的内部结构图。

具体实施方式

[0085] 为了使本发明实现的技术手段、创作特征、达成目的与功效易于明白了解,下面结合具体图示,进一步阐述本发明。

[0086] 本发明提供的政府政策解构方法,可以应用于如图1所示的应用环境中。其中,用户终端101通过网络与服务器102进行通信。其中,用户终端101 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器102可以用独立的服务器或者是多个服务器组成的服务器集群来实现。用户通过用户终端101向服务器102发送信息获取请求,服务器102接收到信息获取请求后,获取政策文件,并对政策文件进行分词

处理,再从分解后的若干个词中逐一找出需概括的关键性词句中的每一个词,使得找出来的这些词构建成概括后的关键性词句,最后将构建后的关键性词句发送至用户终端101。本发明的概括准确率高,有效地提高了工作效率,节省了工作时间,降低了人力成本。

[0087] 参见图2,图中给出的是一种政府政策解构方法,包括以下步骤:

[0088] 步骤S10,确定需概括的关键性词句的平均词数。

[0089] 步骤S20,获取政府网站出台的政策文本。

[0090] 步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词。

[0091] 步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词。

[0092] 步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词。

[0093] 步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。

[0094] 在步骤S10中,参见图3,确定需概括的关键性词句的平均词数,包括以下步骤:

[0095] 步骤S11,获取历史数据库中所有的关键性词句。其中,历史数据库内存储有多个已概括好的政策文件的关键性词句,通过人工方式将经由人工概括好的政策文件的关键性词句输入至历史数据库内进行保存,又或者将每次经过本发明构建后的关键性词句保存至历史数据库内进行保存。

[0096] 步骤S12,对每一关键性词句的词数进行分词处理,使得每一关键性词句被分解成若干个词。在本实施例中,分词处理方法可采用百度LAC分词处理方法。

[0097] 步骤S13,对分词处理后的每一关键性词句的词数进行统计,并对统计后的数据进行平均计算处理,计算所得到的数值为需概括的关键性词句的平均词数。

[0098] 在步骤S20中,获取政府网站出台的政策文本的方式可以为通过互联网直接从政府网站上获取政策文本,又或者通过人工方式将政策文本直接输入至系统内。

[0099] 在步骤S30中,参见图4,对获取到的政策文本进行分词处理,包括以下步骤:

[0100] 步骤S31,对获取到的政策文本进行分段处理,使得所述政策文本被分解成若干个段落。

[0101] 步骤S32,对分解后的每一个段落进行分句处理,使得每一个段落被分解成若干个句子。

[0102] 步骤S33,对分解后的每一个句子进行分词处理,使得每一个句子被分解成若干个词。在本实施例中,分词处理方法可采用百度LAC分词处理方法。

[0103] 在步骤S40中,参见图5,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词,包括以下步骤:

[0104] 步骤S41,从词顺序库内找出第一顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入步骤S42,若匹配失败,则进入步骤 S43。

[0105] 步骤S42,将匹配到的词锁定为需概括的关键性词句的第一顺位的词。

[0106] 步骤S43,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则进入步骤S44。

[0107] 步骤S44,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S42,若匹配失败,则返回步骤S43。

[0108] 在步骤S50中,参见图6,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,包括以下步骤,

[0109] 步骤S51,根据锁定的当前顺位的词从神经元库内找出与之匹配权重最高的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入步骤S52,若匹配失败,则进入步骤S53。

[0110] 步骤S52,将匹配到的词锁定为需概括的关键性词句的下一顺位的词。

[0111] 步骤S53,降低匹配权重,并根据锁定的当前顺位的词从神经元库内找出与之降低后的匹配权重对应的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S54。

[0112] 步骤S54,从词顺序库内找出当前顺位的下一顺位的词,此时该词成为当前顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则进入返回步骤S52,若匹配失败,则进入步骤S55。

[0113] 步骤S55,从近义词库内找出与从词顺序库内找出的当前顺位的词的近义词,并将该近义词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则进入步骤S56。

[0114] 步骤S56,从词顺序库内找出当前顺位的下一顺位的词,并将该词与分解后的若干个词进行逐一匹配,若匹配成功,则返回步骤S52,若匹配失败,则返回步骤S54。

[0115] 参见图7,图中给出的是本发明的政府政策解构方法的具体应用实施例,其包括以下步骤:

[0116] 1. 确定需概括的关键性词句的平均词数。在该实施例中,平均词数为8,则需概括的关键性词句的词数确定为8。

[0117] 2. 获取政府网站出台的政策文本。在该实施例中,该政府文本为“对获得国家级、省级和市级政府质量奖的培育企业,市级财政一次性给予最高500 万元、200万元、100万元的分档奖励。”。

[0118] 3. 对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词。在该实施例中,该政策文本被分解为“(对) (获得) (国家级) (、) (省级) (和) (市级) (政府) (质量奖) (的) (培育) (企业) (,) (市级) (财政) (一次性) (给予) (最高) (500万元) (、) (200万元) (、) (100万元) (的) (分档) (奖励) (。)”。

[0119] 4. 通过词顺序库和近义词库从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词。

[0120] 5. 根据锁定的当前顺位的词,通过词顺序库和近义词库从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词。

[0121] 6. 重复执行上述步骤5,直至需概括的关键性词句的词数达到平均词数。

[0122] 参见图8,本发明的政府政策解构方法还包括步骤S70,对概括形成的关键性词句

进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,返回步骤S10,对关键性词句进行重新构建。

[0123] 参见图9,本发明的政府政策解构方法还包括步骤S80,对语法判断处理后的关键性词句进行人工校验,若人工校验成功,则将概括形成的关键性词句推送至目标者,若人工校验失败,则返回步骤S10,对关键性词句进行重新构建。

[0124] 参见图10,本发明的政府政策解构方法还包括步骤S90,对人工校验成功的关键性词句进行自我学习处理,以提高概括的准确性。

[0125] 在步骤S90中,参见图11,对人工校验成功的关键性词句进行自我学习处理,包括以下步骤:

[0126] 步骤S91,对人工校验成功的关键性词句进行分词处理,使得关键性词句被分解为若干个词。在本实施例中,分词处理方法可采用百度LAC分词处理方法。

[0127] 步骤S92,建立每相邻两个词之间的匹配权重,并将建立后的匹配权重保存至神经元库内。

[0128] 步骤S93,对分解后的每个词进行词性识别,并将识别结果保存至词性库内。

[0129] 步骤S94,从互联网上获取分解后的每个词的近似词,并将获取到的这些近义词保存至近义词库内。

[0130] 步骤S95,对分解后的每个词进行优先级排序处理,并将优先级排序结果保存至词顺序库内。

[0131] 参见图12,图中给出的是一种政府政策关键性信息构建装置,包括平均词数计算模块100、政策文本获取模块200、分词处理模块300、第一词句匹配模块400、第二词句匹配模块500、语法判断处理模块600以及自我学习模块700。

[0132] 平均词数计算模块100用于确定需概括的关键性词句的平均词数。

[0133] 政策文本获取模块200用于获取政府网站出台的政策文本。

[0134] 分词处理模块300用于对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词。

[0135] 第一词句匹配模块400用于从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词。

[0136] 第二词句匹配模块500用于根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词,如此重复,直至需概括的关键性词句的词数达到平均词数。

[0137] 语法判断处理模块600用于对概括形成的关键性词句进行语法判断处理,若语法判断正确,则表明概括形成的关键性词句符合语法规则,若语法判断错误,则表明概括形成的关键性词句不符合语法规则,对关键性词句进行重新构建。

[0138] 自我学习模块700用于对概括形成的关键性词句进行自我学习处理,其目的是为了提提高概括准确率。具体地,自我学习模块700先对人工校验成功的关键性词句进行分词处理,使得关键性词句被分解为若干个词。然后,建立每相邻两个词之间的匹配权重,并将建立后的匹配权重保存至神经元库内。接着,对分解后的每个词进行词性识别,并将识别结果

保存至词性库内。继而,从互联网上获取分解后的每个词的近似词,并将获取到的这些近义词保存至近义词库内。最后,对分解后的每个词进行优先级排序处理,并将优先级排序结果保存至词顺序库内。

[0139] 本发明的政府政策关键性信息构建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。

[0140] 本发明还提供了一种用于实现上述政府政策解构方法的计算机设备,该计算机设备可以是服务器,其内部结构图可以如图13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、网络接口和数据库。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、计算机程序和数据库。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。该计算机设备的数据库用于存储用户信息、记录信息和文件等数据。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上述的政府政策解构方法。

[0141] 本领域技术人员可以理解,图13中示出的结构,仅仅是与本技术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。

[0142] 具体地,本发明的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,处理器执行计算机程序时实现以下步骤:

[0143] 步骤S10,确定需概括的关键性词句的平均词数;

[0144] 步骤S20,获取政府网站出台的政策文本;

[0145] 步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;

[0146] 步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;

[0147] 步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;

[0148] 步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。

[0149] 本发明还提供了一种用于实现上述政府政策解构方法的计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以下步骤:

[0150] 步骤S10,确定需概括的关键性词句的平均词数;

[0151] 步骤S20,获取政府网站出台的政策文本;

[0152] 步骤S30,对获取到的政策文本进行分词处理,以使得所述政策文本被分解成若干个词;

[0153] 步骤S40,从分解后的若干个词中找出作为需概括的关键性词句的第一顺位的词;

[0154] 步骤S50,根据锁定的当前顺位的词从分解后的若干个词中找出与之匹配权重最高的词,并将该词锁定为需概括的关键性词句的下一顺位的词,此时该词成为需概括的关键性词句的当前顺位的词;

[0155] 步骤S60,重复执行步骤S50,直至需概括的关键性词句的词数达到平均词数。

[0156] 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取存储介质中,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。

[0157] 以上显示和描述了本发明的基本原理和主要特征和本发明的优点。本行业的技术人员应该了解,本发明不受上述实施例的限制,上述实施例和说明书中描述的只是说明本发明的原理,在不脱离本发明精神和范围的前提下,本发明还会有各种变化和改进,这些变化和改进都落入要求保护的本发明范围内。本发明要求保护范围由所附的权利要求书及其等效物界定。

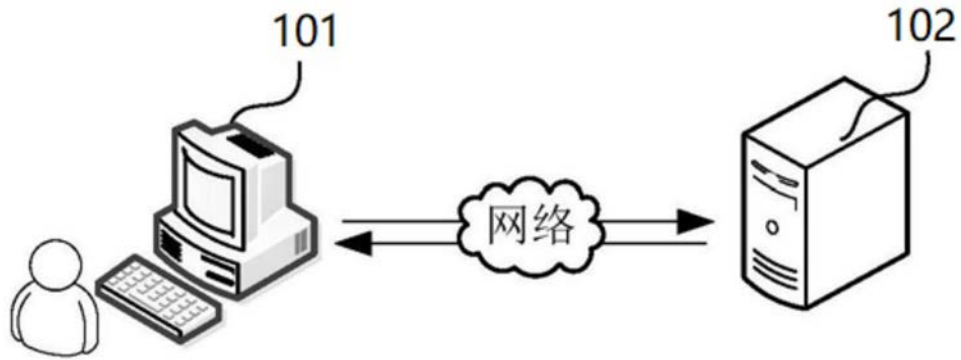


图1

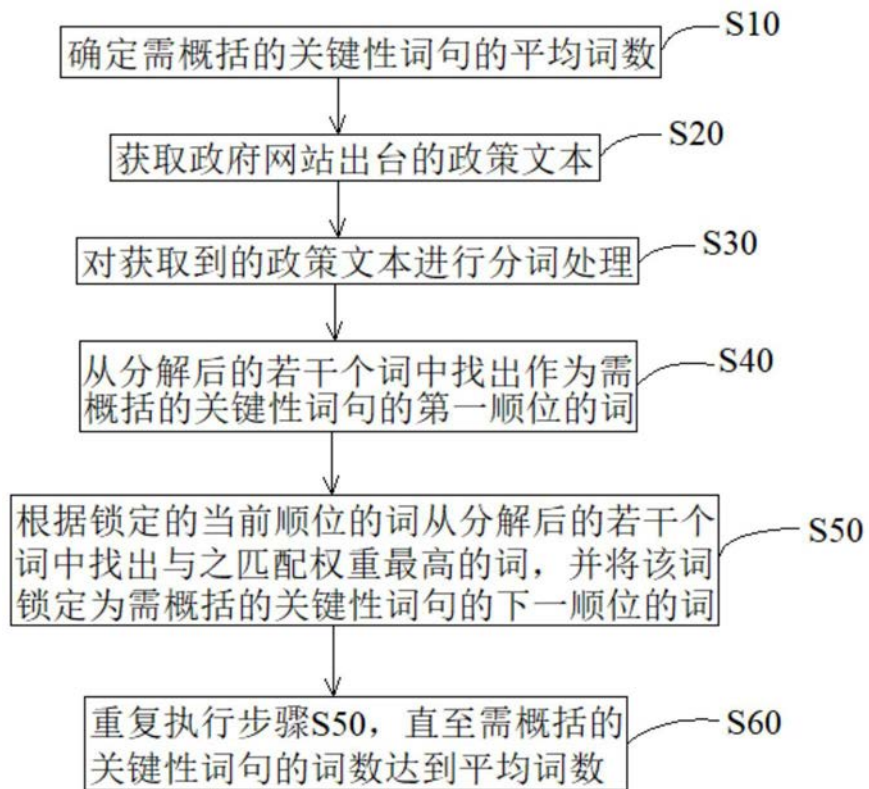


图2

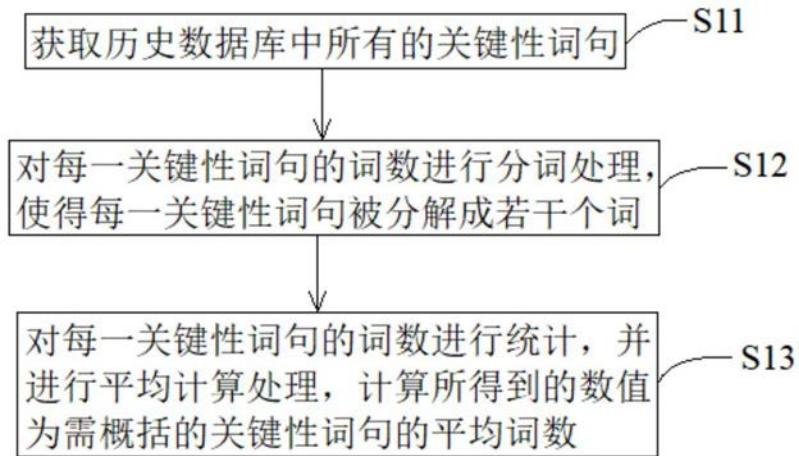


图3

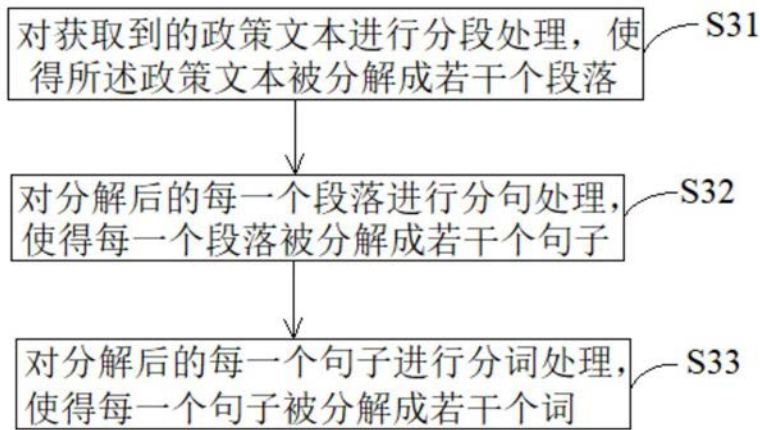


图4

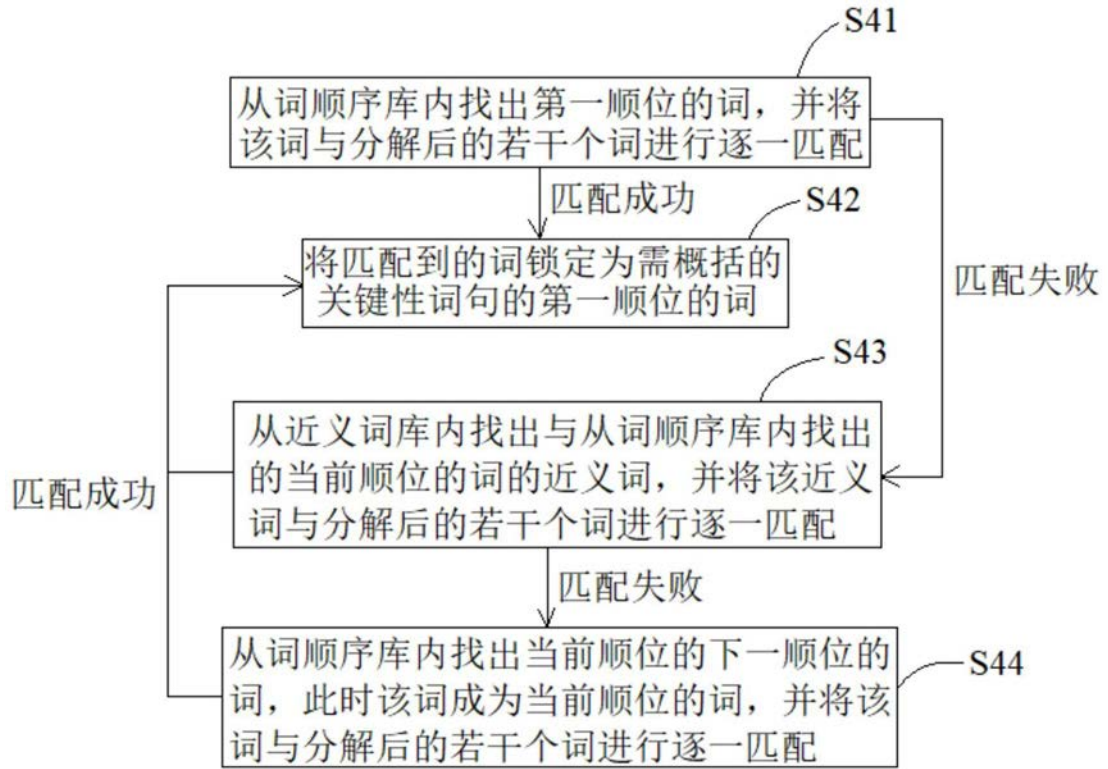


图5

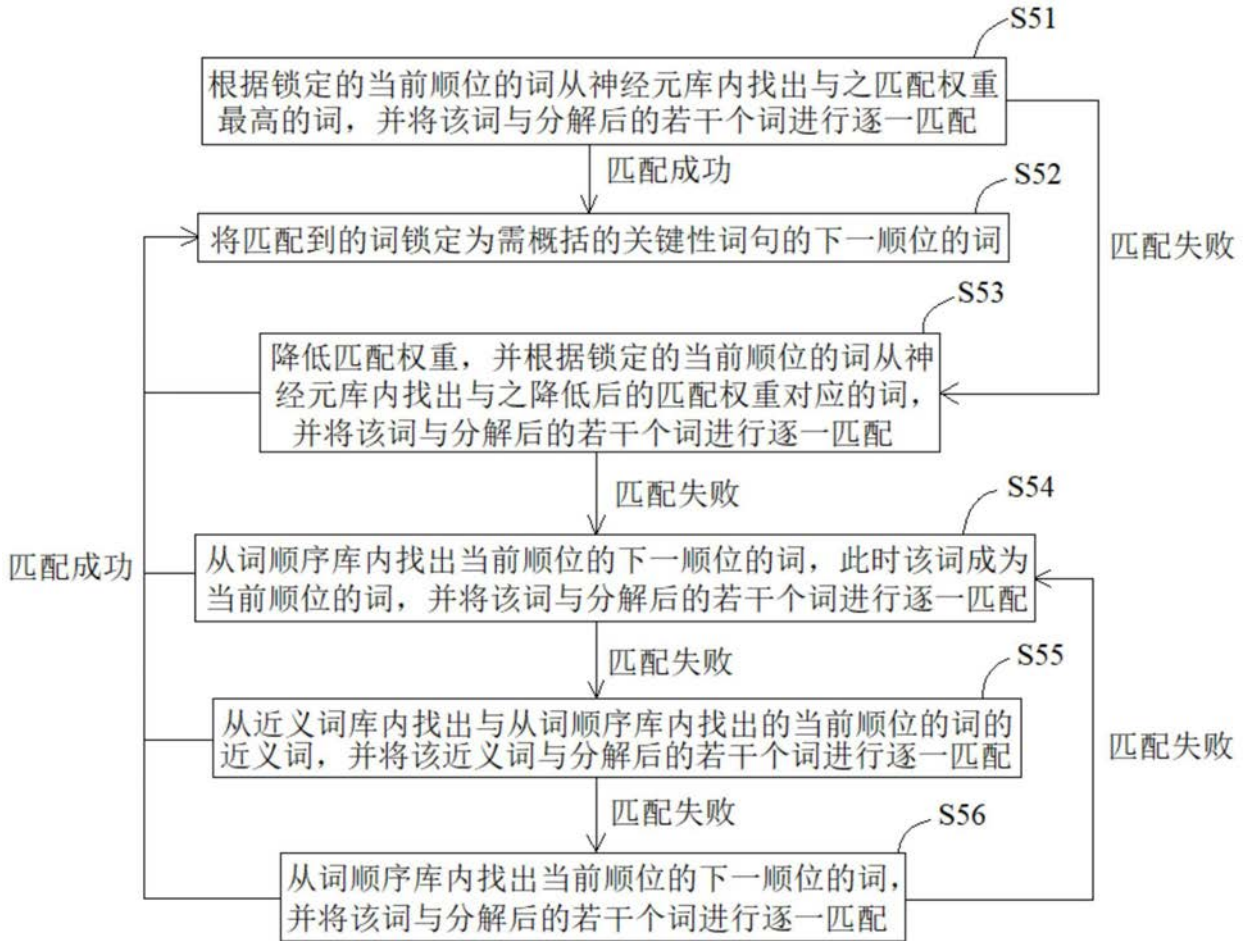


图6

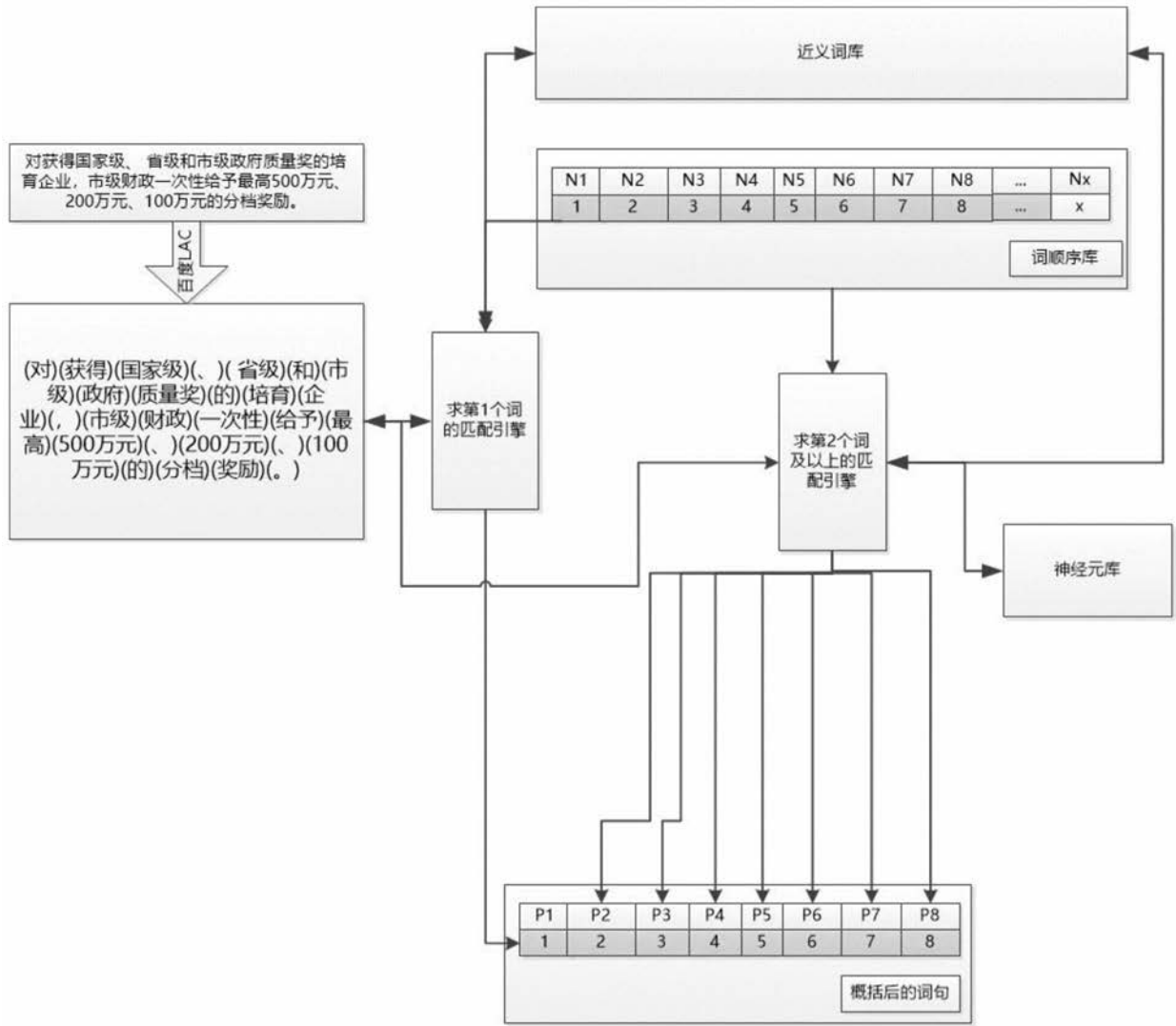


图7

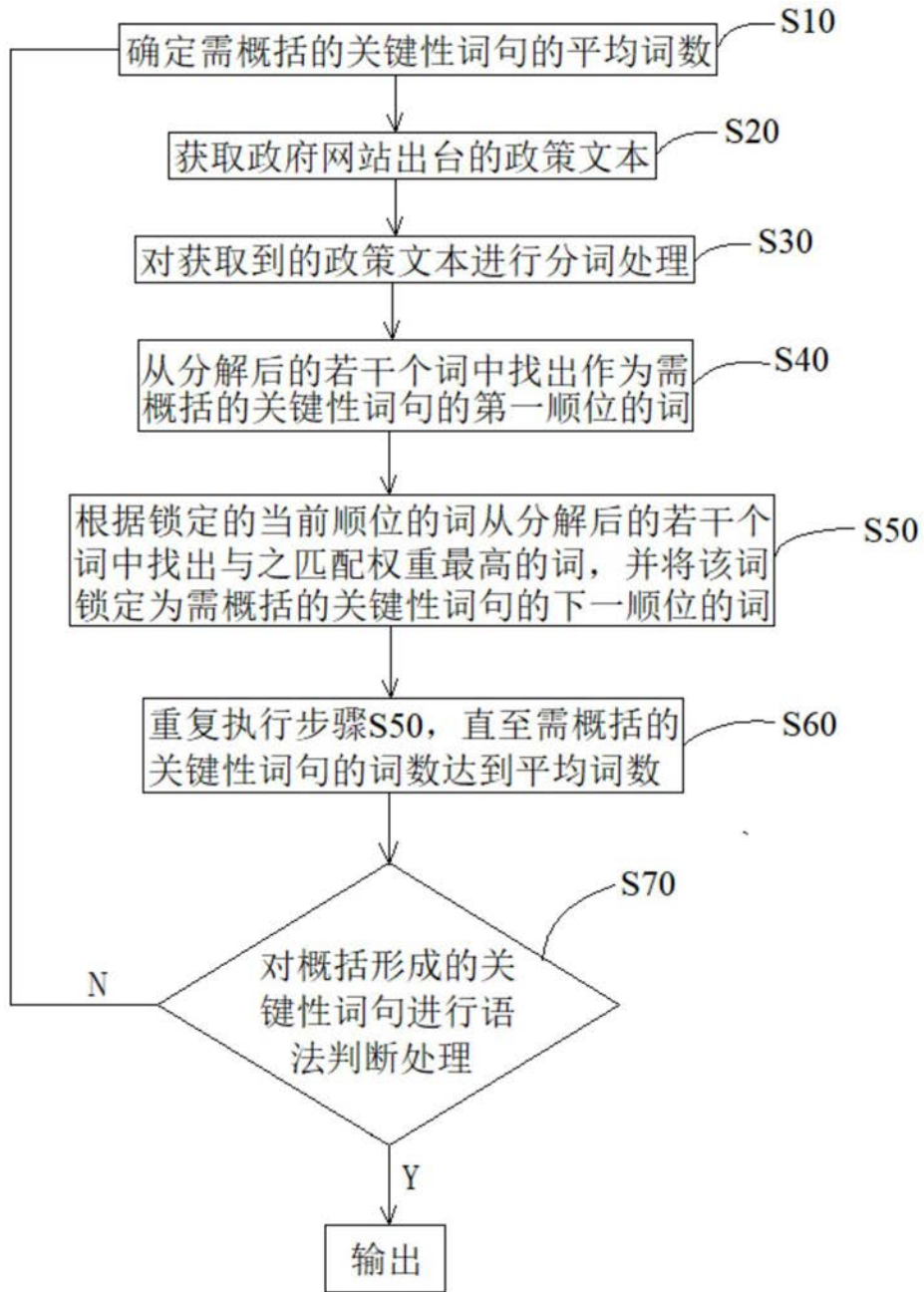


图8

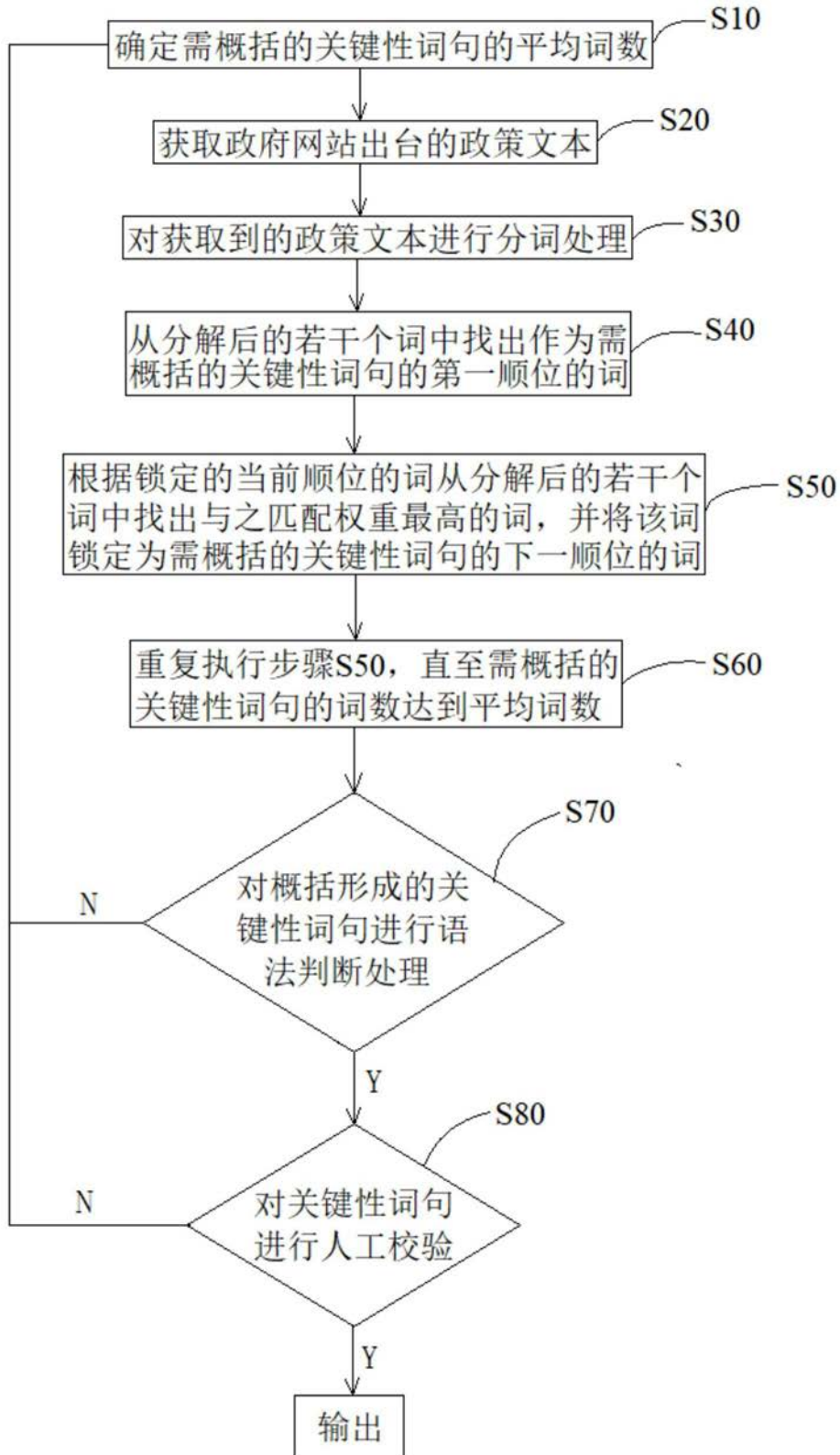


图9

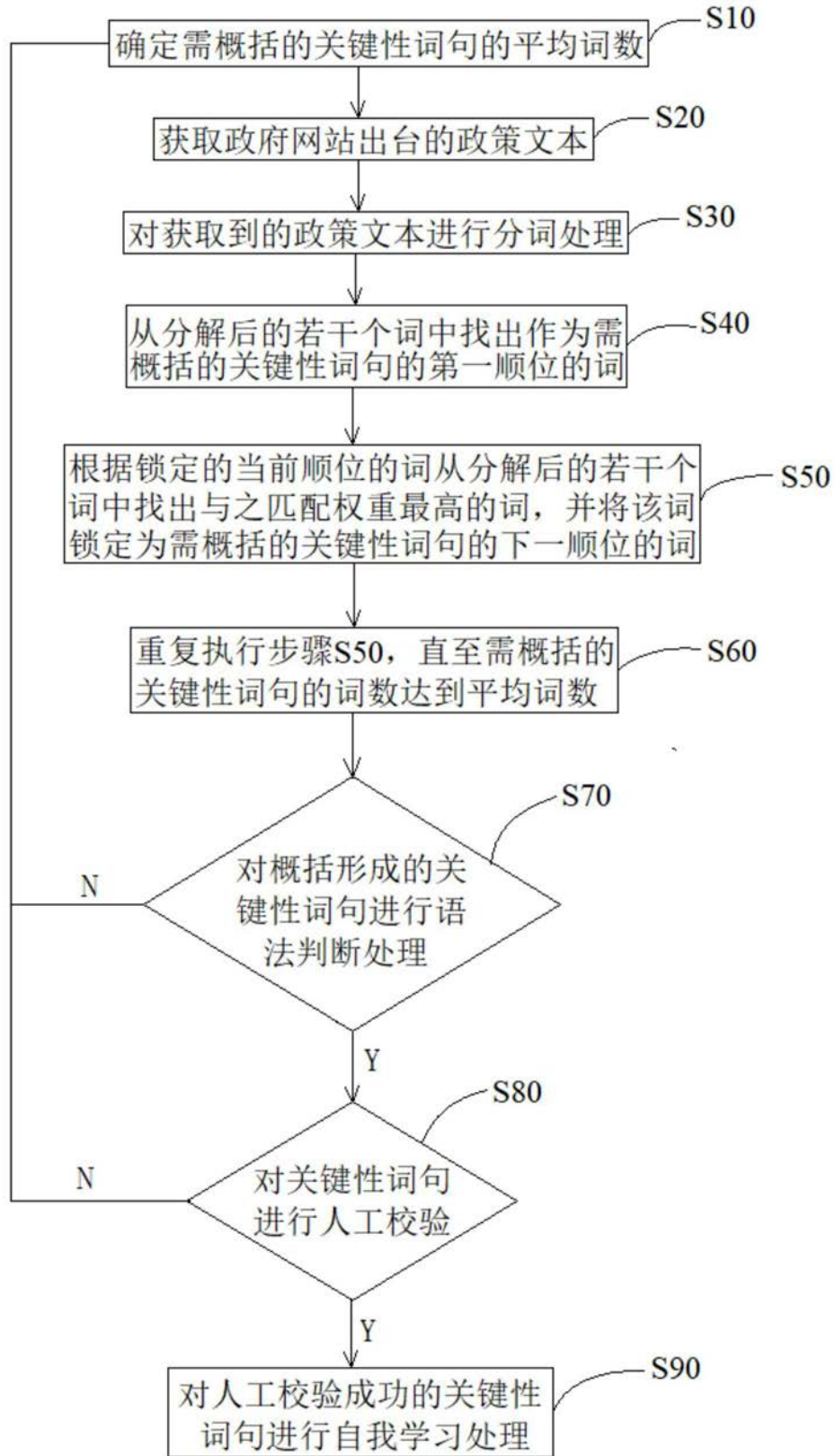


图10

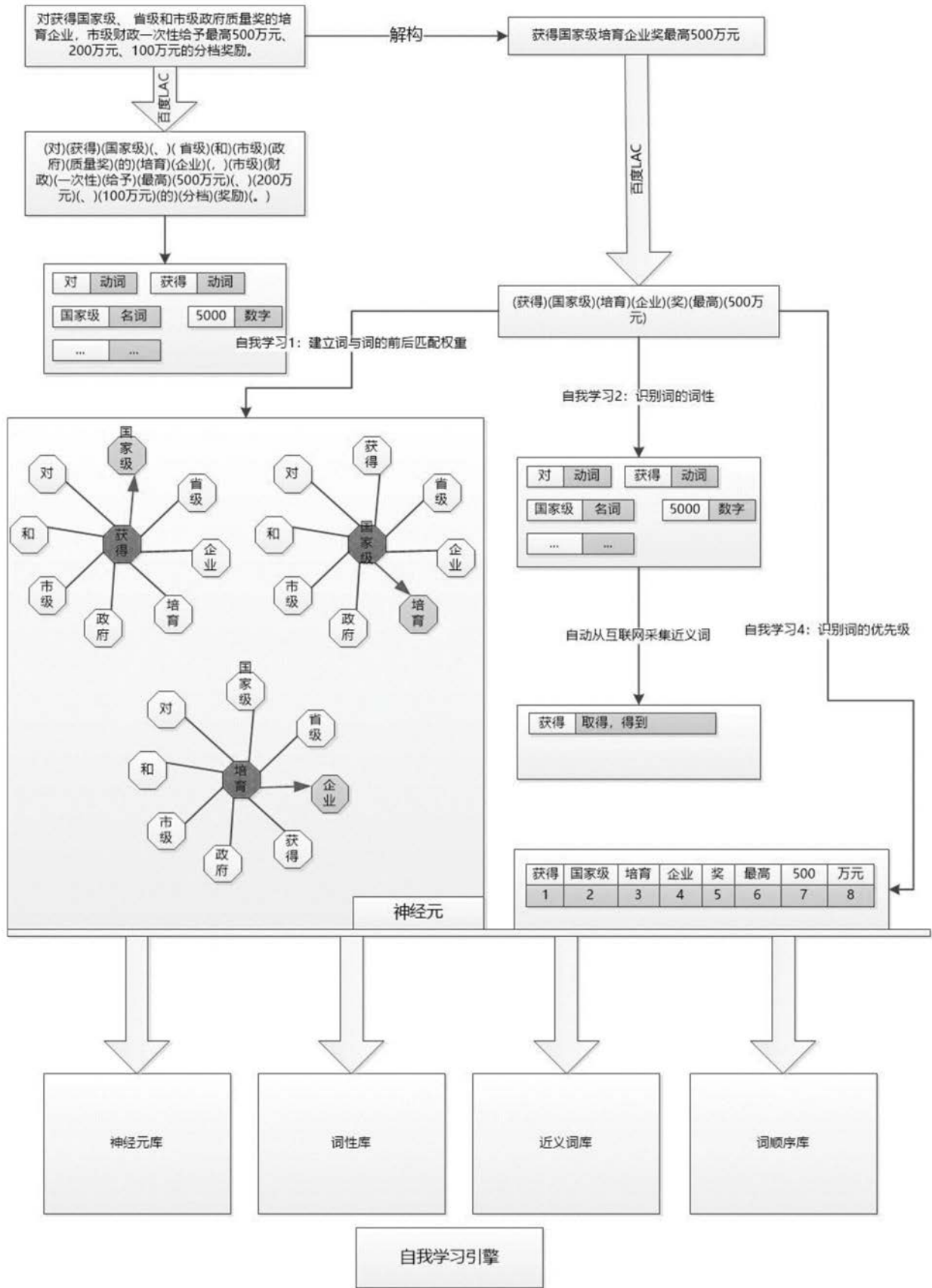


图11

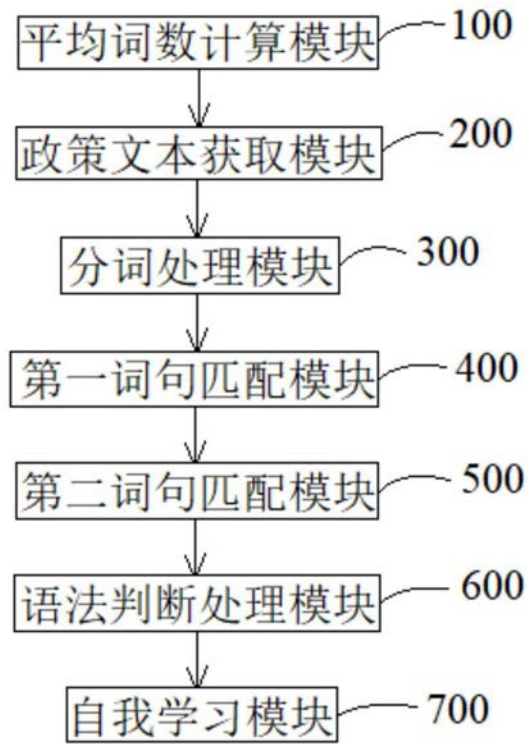


图12

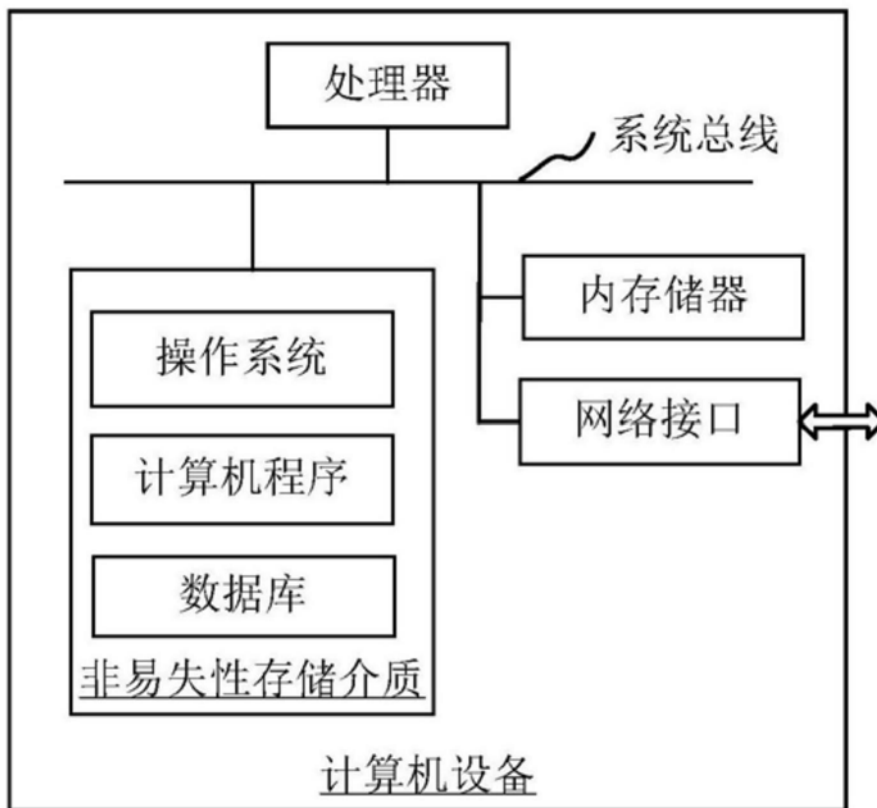


图13