



(12)发明专利

(10)授权公告号 CN 105068995 B

(45)授权公告日 2018.05.29

(21)申请号 201510510604.5

CN 104361127 A,2015.02.18,

(22)申请日 2015.08.19

CN 1952928 A,2007.04.25,

(65)同一申请的已公布的文献号

US 5794050 A,1998.08.11,

申请公布号 CN 105068995 A

EP 1395914 A1,2004.03.10,

CN 104142917 A,2014.11.12,

(43)申请公布日 2015.11.18

陈庚等.基于问句语义表征的中文问句相似度计算方法.《北京理工大学学报》.2007,第27卷(第12期),

(73)专利权人 刘战雄

于甜甜.基于语义树的语句相似度和相关度在问答系统中的研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2014,第2014年卷(第09期),

地址 065201 河北省廊坊市三河市燕郊开

发区福成五期25-3-905

(72)发明人 刘战雄

(74)专利代理机构 北京品源专利代理有限公司

11332

代理人 孟金喆 胡彬

马玉慧等.基于语义句模的语义理解方法研究.《计算机技术与发展》.2012,第22卷(第10期),

审查员 宿渊源

(51)Int.Cl.

G06F 17/27(2006.01)

(56)对比文件

CN 104657463 A,2015.05.27,

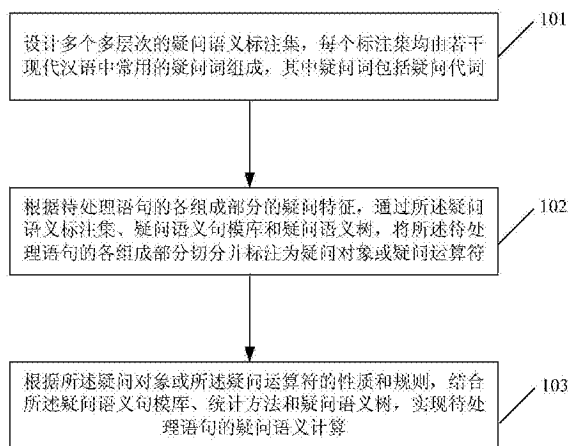
权利要求书3页 说明书13页 附图5页

(54)发明名称

一种基于疑问语义的自然语言语义计算的方法及装置

(57)摘要

本发明实施例公开了一种基于疑问语义的自然语言语义计算方法及装置,设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;根据所述疑问对象或所述疑问运算符的性质和规则,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,作为一个基础的语义计算方法,本发明实施例可以有效的解决常见的自然语言处理问题,尤其是在语句的切分标注、自然语言搜索、机器翻译、人机问答等领域具有较高的使用价值。



1. 一种基于疑问语义的自然语言语义计算的方法,其特征在于,所述方法包括:

设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;

根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;利用疑问对象性质、运算符性质以及运算规则进行处理,进而建立所述疑问语义树,所述疑问语义树的节点为字符块,所述疑问语义树的边为标注符号,所述疑问语义树可以分解与合并;

通过统计疑问语义树及疑问语义子树,统计出相应的疑问语义句模,进而建立疑问语义句模库,所述疑问语义句模库用来驱动语义规则;

根据所述疑问对象或所述疑问运算符的性质和规则,结合所述疑问语义句模库、统计方法和疑问语义树,实现待处理语句的疑问语义计算;

所述方法还包括:

将搜索字符通过预先设置的算法划分为预先设置的疑问对象;

根据划分后的疑问代词搜索预先存储的字符;

若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

2. 根据权利要求1所述的方法,其特征在于,所述设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词,包括:

根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词。

3. 根据权利要求1所述的方法,其特征在于,所述根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符,包括:

若所述语义范畴为人,则与人对应的疑问代词为“谁”;

若所述语义范畴为事物,则与事物对应的疑问代词为“什么”;

若所述语义范畴为动作,则与所述动作对应的疑问代词为“怎么”,与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何;

若所述语义范畴为时间,则与所述时间对应的疑问代词为几时,与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱;

若所述语义范畴为地点场所,则与所述地点场所对应的疑问代词为哪里,与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿;

若所述语义范畴为数目数量,则与所述数目数量对应的疑问代词为多少,与所述多少等同的其他疑问代词为几和多;

若所述语义范畴为虚词,则将所述语义范畴切分并标注为疑问运算符。

4. 根据权利要求1至3任意一项所述的方法,其特征在于,所述方法还包括:

接收用户输入的搜索字符;

根据所述搜索字符和相似度计算获取预先存储的字符模型;

根据所述预先存储的字符模块划分为预先设置的疑问对象;

根据划分后的疑问代词搜索预先存储的字符；

若划分后的疑问代词和预先存储的字符对应，则显示与所述预先存储的字符对应的划分前的待处理的字符。

5. 一种基于疑问语义的自然语言语义计算的装置，其特征在于，所述装置包括：

设计模块，用于设计多个多层次的疑问语义标注集，每个标注集均由若干现代汉语中常用的疑问词组成，其中疑问词包括疑问代词；

标注模块，用于根据待处理语句的各组成部分的疑问特征，通过所述疑问语义标注集、疑问语义句模库和疑问语义树，将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符；利用疑问对象性质、运算符性质以及运算规则进行处理，进而建立所述疑问语义树，所述疑问语义树的节点为字符块，所述疑问语义树的边为标注符号，所述疑问语义树可以分解与合并；通过统计疑问语义树及疑问语义子树，统计出相应的疑问语义句模，进而建立疑问语义句模库，所述疑问语义句模库用来驱动语义规则；

计算模块，用于根据所述疑问对象或所述疑问运算符的性质和规则，结合所述疑问语义句模库、统计方法和疑问语义树，实现待处理语句的疑问语义计算；

所述装置还包括：

第一划分模块，用于将搜索字符通过预先设置的算法划分为预先设置的疑问对象；

第一搜索模块，用于根据划分后的疑问代词搜索预先存储的字符；

第一显示模块，用于若划分后的疑问代词和预先存储的字符对应，则显示与所述预先存储的字符对应的划分前的待处理的字符。

6. 根据权利要求5所述的装置，其特征在于，所述设计模块，包括：

设计单元，用于根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集，每个标注集均由若干现代汉语中常用的疑问词组成，其中疑问词包括疑问代词。

7. 根据权利要求5所述的装置，其特征在于，所述标注模块，包括：

第一标注单元，用于若所述语义范畴为人，则与人对应的疑问代词为“谁”；

第二标注单元，用于若所述语义范畴为事物，则与事物对应的疑问代词为“什么”；

第三标注单元，用于若所述语义范畴为动作，则与所述动作对应的疑问代词为“怎么”，与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何；

第四标注单元，用于若所述语义范畴为时间，则与所述时间对应的疑问代词为几时，与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱；

第五标注单元，用于若所述语义范畴为地点场所，则与所述地点场所对应的疑问代词为哪里，与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿；

第六标注单元，用于若所述语义范畴为数目数量，则与所述数目数量对应的疑问代词为多少，与所述多少等同的其他疑问代词为几和多；

第七标注单元，用于若所述语义范畴为虚词，则将所述语义范畴切分并标注为疑问运算符。

8. 根据权利要求5至7任意一项所述的装置，其特征在于，所述装置还包括：

接收模块，用于接收用户输入的搜索字符；

获取模块，用于根据所述搜索字符和相似度计算获取预先存储的字符模型；

第二划分模块，用于根据所述预先存储的字符模块划分为预先设置的疑问对象；

第二搜索模块,用于根据划分后的疑问代词搜索预先存储的字符;

第二显示模块,用于若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

一种基于疑问语义的自然语言语义计算的方法及装置

技术领域

[0001] 本发明实施例涉及信息处理的技术领域,尤其涉及一种基于疑问语义的自然语言语义计算的方法及装置。

背景技术

[0002] 自然语言处理是研究人人、人机交际中语言问题的一门学科。在自然语言处理中,语义计算是通过计算机解释自然语言中各级单位即字、词素、词、词组、短语、句子、句群、段落、篇章等的意义,所关心的重点是这个语言单位到底说了什么。在目前技术中主要有:义素分析法、语义场、语义网络、Montague语法、优选语义学、概念依存理论、意义-文本理论等方法。

[0003] 在目前技术中,语义计算的主要的缺陷表现为两方面:一方面侧重于利用统计方法进行字符运算,很少或者没有涉及语义问题;另一方面失之于语义概念抽象或规则复杂,利用计算机技术难以实现或算法复杂度高,缺乏实用性。

发明内容

[0004] 本发明实施例的目的在于提出一种基于疑问语义的自然语言语义计算的方法及装置,旨在解决如何建立简单易懂的语义划分规则对自然语言进行处理的问题。

[0005] 为达此目的,本发明实施例采用以下技术方案:

[0006] 设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;

[0007] 根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;

[0008] 根据所述疑问对象或所述疑问运算符的性质和规则,结合所述疑问语义句模库、统计方法和疑问语义树,实现待处理语句的疑问语义计算。

[0009] 优选地,所述设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词,包括:

[0010] 根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词。

[0011] 优选地,所述根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符,包括:

[0012] 若所述语义范畴为人,则与人对应的疑问代词为“谁”;

[0013] 若所述语义范畴为事物,则与事物对应的疑问代词为“什么”;

[0014] 若所述语义范畴为动作,则与所述事物对应的疑问代词为“怎么”,与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何;

- [0015] 若所述语义范畴为时间,则与所述时间对应的疑问代词为几时,与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱;
- [0016] 若所述语义范畴为地点场所,则与所述地点场所对应的疑问代词为哪里,与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿;
- [0017] 若所述语义范畴为数目数量,则与所述数目数量对应的疑问代词为多少,与所述多少等同的其他疑问代词为几和多;
- [0018] 若所述语义范畴为虚词,则将所述语义范畴切分并标注为疑问运算符。
- [0019] 优选地,所述方法还包括:
- [0020] 将搜索字符通过预先设置的算法划分为预先设置的疑问对象;
- [0021] 根据划分后的疑问代词搜索预先存储的字符;
- [0022] 若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。
- [0023] 优选地,所述方法还包括:
- [0024] 接收用户输入的搜索字符;
- [0025] 根据所述搜索字符和相似度计算获取预先存储的字符模型;
- [0026] 根据所述预先存储的字符模块划分为预先设置的疑问对象;
- [0027] 根据划分后的疑问代词搜索预先存储的字符;
- [0028] 若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。
- [0029] 一种基于疑问语义的自然语言语义计算的装置,所述装置包括:
- [0030] 设计模块,用于设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;
- [0031] 标注模块,用于根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;
- [0032] 计算模块,用于根据所述疑问对象或所述疑问运算符的性质和规则,结合所述疑问语义句模库、统计方法和疑问语义树,实现待处理语句的疑问语义计算。
- [0033] 优选地,所述设计模块,包括:
- [0034] 设计单元,用于根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词。
- [0035] 优选地,所述标注模块,包括:
- [0036] 第一标注单元,用于若所述语义范畴为人,则与人对应的疑问代词为“谁”;
- [0037] 第二标注单元,用于若所述语义范畴为事物,则与事物对应的疑问代词为“什么”;
- [0038] 第三标注单元,用于若所述语义范畴为动作,则与所述事物对应的疑问代词为“怎么”,与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何;
- [0039] 第四标注单元,用于若所述语义范畴为时间,则与所述时间对应的疑问代词为几时,与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱;
- [0040] 第五标注单元,用于若所述语义范畴为地点场所,则与所述地点场所对应的疑问

代词为哪里,与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿;

[0041] 第六标注单元,用于若所述语义范畴为数目数量,则与所述数目数量对应的疑问代词为多少,与所述多少等同的其他疑问代词为几和多;

[0042] 第七标注单元,用于若所述语义范畴为虚词,则将所述语义范畴切分并标注为疑问运算符。

[0043] 优选地,所述装置还包括:

[0044] 第一划分模块,用于将搜索字符通过预先设置的算法划分为预先设置的疑问对象;

[0045] 第一搜索模块,用于根据划分后的疑问代词搜索预先存储的字符;

[0046] 第一显示模块,用于若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

[0047] 优选地,接收模块,用于接收用户输入的搜索字符;

[0048] 获取模块,用于根据所述搜索字符和相似度计算获取预先存储的字符模型;

[0049] 第二划分模块,用于根据所述预先存储的字符模块划分为预先设置的疑问对象;

[0050] 第二搜索模块,用于根据划分后的疑问代词搜索预先存储的字符;

[0051] 第二显示模块,用于若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

[0052] 本发明实施例通过设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;根据所述疑问对象或所述疑问运算符的性质和规则,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,实现待处理语句的疑问语义计算,作为一个基础的语义计算方法,本发明实施例可以有效的解决常见的自然语言处理问题,尤其是在语句的切分标注、自然语言搜索、机器翻译、人机问答等领域具有较高的使用价值。

附图说明

[0053] 图1是本发明实施例基于疑问语义的自然语言语义计算的方法第一实施例的流程示意图;

[0054] 图2是本发明实施例基于疑问语义的自然语言语义计算的方法第二实施例的流程示意图;

[0055] 图3是本发明实施例基于疑问语义的自然语言语义计算的方法第三实施例的流程示意图;

[0056] 图4是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图;

[0057] 图5是本发明实施例设计模块401的功能模块示意图;

[0058] 图6是本发明实施例标注模块402的功能模块示意图;

[0059] 图7是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图;

[0060] 图8是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图。

具体实施方式

[0061] 下面结合附图和实施例对本发明实施例作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释本发明实施例,而非对本发明实施例的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本发明实施例相关的部分而非全部结构。

[0062] 实施例一

[0063] 参考图1,图1是本发明实施例基于疑问语义的自然语言语义计算的方法第一实施例的流程示意图。

[0064] 在实施例一中,所述基于疑问语义的自然语言语义计算的方法包括:

[0065] 步骤101,设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;

[0066] 优选地,所述设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词,包括:

[0067] 根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词。

[0068] 步骤102,根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;

[0069] 优选地,所述根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符,包括:

[0070] 若所述语义范畴为人,则与人对应的疑问代词为“谁”;

[0071] 若所述语义范畴为事物,则与事物对应的疑问代词为“什么”;

[0072] 若所述语义范畴为动作,则与所述事物对应的疑问代词为“怎么”,与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何;

[0073] 若所述语义范畴为时间,则与所述时间对应的疑问代词为几时,与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱;

[0074] 若所述语义范畴为地点场所,则与所述地点场所对应的疑问代词为哪里,与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿;

[0075] 若所述语义范畴为数目数量,则与所述数目数量对应的疑问代词为多少,与所述多少等同的其他疑问代词为几和多;

[0076] 若所述语义范畴为虚词,则将所述语义范畴切分并标注为疑问运算符。

[0077] 具体的,现代汉语中的疑问代词是一个相对封闭的词类。现代汉语中的疑问代词从使用频率上来看,常见的疑问代词有:什么、怎么、谁、几、哪、为什么、多少、哪儿、怎么样、何、哪里、如何、多、干吗;非常用的疑问代词:怎样、怎、怎么着、为何、怎的、几时、哪会儿、多会儿、多咱。

[0078] 鉴于以下两点:

[0079] 一个是某些方言中的疑问代词,如:啥、弄啥、干嘛、干吗、咋办、咋整等,其基本疑问语义可用现代汉语中的疑问代词来替换;另一个是组合语义的疑问代词或短语,如:什么时候、什么地方、什么人、什么事、什么东西、什么样子、什么样的、什么数目、什么数量、什么高度、什么重量、什么程度、什么情况、何人、何时、何处、何物、为何等,其基本疑问语义均能为现代汉语中的疑问代词直接或组合替代,因此不再单独罗列到疑问代词集。

[0080] 将疑问代词的基本疑问语义限定在范畴层面进行分析后,可知其能对:人、事物、时间、处所、数量、方式、性状、原因等范畴内的对象进行疑问。在疑问代词的分类上,我们采取求同存异(动态的抓大放小)的权衡策略,即寻求其在疑问范畴层面上的一致之处(权值大),而忽略其语义或用法上的细微差异之处(权值小)。对于疑问代词,当其疑问范畴一致时,我们将其视为一类。如:“哪儿、哪里”可对表达地点类的字符块提出疑问,因此视为同类。

[0081] 对于“什么”这一特殊疑问代词,由于其可以对任何字符块提出“未知”疑问,“未知”对于了解字符块意义的贡献有限,疑问对象十分宽泛,因此我们这里采取的是缩小其疑问范畴,将其作为对表达事物类语义的字符块进行提问的疑问代词。鉴于其能跟多类字符块组合起来表达某一特定疑问范畴,1) 其组合语义可被某单一疑问代词等价替换,如:“什么地方”可被“哪里”替换;2) 涉及范畴内对象的深层次语义分析,如:“什么高度、什么宽度、什么长度”等,对于我们来说,在疑问范畴层面上属于数量范畴内的深层次语义,因此在范畴层面暂不处理;3) 要么涉及范畴内对象的综合语义分析,如:什么情况、什么原因等,此是疑问对象多个范畴层面的综合语义结果,因此在范畴层面暂不处理。

[0082] 为了利用计算机理解并处理语义,在限于范畴层面考虑疑问代词的语义时,我们从疑问语义角度入手,利用疑问代词集中的疑问代词标注疑问对象(或它的各个义项)所归属的语义范畴,即:可用哪种疑问词对疑问对象进行提问。

[0083] 根据疑问代词的疑问范畴来设计多种疑问语义标注集,以便对疑问对象进行标注,例如以下为所设计的多个多层次疑问语义标注集中具有代表性的一个:

[0084] 形式化表示为:Y={谁,什么,怎样,几时,哪里,多少,……}

[0085] 具体的,参考如下表1:

[0086]

疑问对象的语义范畴	用来标注的疑问代词	标注符号
人	谁	
事物(包含性质)	什么	
动作(包含方式)	怎么,等价的有:怎么样、怎样、怎、怎么着、怎的、如何	
时间	几时,等价的有:多会儿、哪会儿、多咱	
地点方所	哪里,等价的有:哪、哪儿	
数目数量	多少,等价的有:几、多	
无实在意义的字符块	疑问运算符	

[0087] 表1

[0088] 在自然语言处理中,语义计算是通过计算机解释自然语言中各级单位的意义,如:字、词、词组、短语、句子、句群、段落、篇章等的意义。为了处理方便,我们假定只考虑句子及组成句子的各级单位的意义。在各级单位中,假设一个有实在意义的字符块,它的意义或某

个义项的意义一定属于某一或某些范畴,并且能被某一个或某些疑问代词所疑问时,称这个字符块为疑问对象。对于一些没有实在意义或无法被疑问的字符块,我们称之为疑问运算符。如:一些常见的实词为疑问对象,而一些常见的虚词为疑问运算符。每个疑问对象都作为一个疑问点,可以用于检索、人机问答以及机器翻译。

[0089] 根据疑问对象的特征或称为性质、属性将其分为若干类别,并制定若干规则。

[0090] 疑问对象的若干属性包含但不局限于:

[0091] 疑问对象或其义项所归属的疑问范畴,即用哪些疑问代词标注该疑问对象;

[0092] 疑问对象与疑问对象的搭配属性;

[0093] 疑问对象与疑问运算符的搭配属性;

[0094] 疑问对象可支配的疑问对象数目(分为一元、二元、三元等);

[0095] 疑问对象的作用范围与运算方向;

[0096] 同类疑问对象间、非同类疑问对象间的合并运算;

[0097] 疑问对象间的语义侧重点;

[0098] 疑问对象的若干运算规则(包含但不局限于):

[0099] 分解运算;

[0100] 复合疑问对象被分解为若干疑问对象;

[0101] 合并运算;

[0102] 多个疑问对象合并为一个疑问对象;

[0103] 顺序变换运算;

[0104] 某些疑问对象的顺序能够改变次序而保持等价语义;

[0105] 分解运算;

[0106] 在某些疑问对象上实施递归的疑问语义处理。

[0107] 具体的,在句子及组成句子的各级单位中,存在非可被疑问的部分或某个义项即非疑问对象,对于这一部分可分为两类:

[0108] 多数虚词,一般不作为疑问对象,在本发明中,称之为疑问运算符;

[0109] 标点符号,由于其数量、用法有限,暂不考虑或特殊处理;

[0110] 对于疑问运算符,根据其特征或称为性质分为若干类别,并根据疑问运算符的特征,制定出若干规则;

[0111] 疑问运算符的若干属性包含但不局限于:

[0112] 疑问运算符可支配的疑问对象;

[0113] 根据运算符可支配的疑问对象数目(分为一元、二元、三元等);

[0114] 运算符的作用范围;

[0115] 运算符的运算方向从左到右,还是右到左。如“被”和“把”的一类语义句子中如考虑到方向问题,则认为:谁被什么怎么了,等价于:什么把谁怎么了。

[0116] 疑问运算符的若干运算规则包含但不局限于:

[0117] 分解运算;

[0118] 疑问对象被分解为疑问对象和运算符。

[0119] 例如:疑问对象:“我和你”可以被“谁”疑问,同时作为一个复合的疑问对象,其可以分解为:疑问对象:“我”、“你”,以及疑问运算符:“和”。

- [0120] 合并运算：
- [0121] 疑问对象与运算符合并为新的疑问对象。
- [0122] 例如：同上例，疑问对象：“我”、“你”，以及疑问运算符：“和”。
- [0123] 我/谁和/疑问运算符你/谁
- [0124] 合并运算后：
- [0125] 我和你/谁
- [0126] 顺序变换运算：
- [0127] 某些疑问对象的顺序能够改变次序而保持等价语义。
- [0128] 例如：同上例，疑问对象：“我”、“你”，以及疑问运算符：“和”。
- [0129] 我和你/谁
- [0130] 顺序变换后，语义保持不变：
- [0131] 你和我/谁
- [0132] 递归运算：
- [0133] 在某些疑问对象上实施递归的疑问语义处理。
- [0134] 步骤103，根据所述疑问对象或所述疑问运算符的性质和规则，结合所述疑问语义句模库、统计方法和疑问语义树，实现待处理语句的疑问语义计算。
- [0135] 具体的，在对句子的字符块进行标注后，我们利用疑问对象性质、运算符性质以及运算规则对其进行处理，进而建立起疑问语义树。将句子中的疑问对象作为疑问点，利用疑问点可以用来回答相应的搜索、翻译或人机对话。疑问点对应疑问语义树中的节点。对于疑问语义树，我们用其节点表示字符块，用边表示标注符号。在这里，我们可以通过分解与合并来操作疑问语义树，一个是将未标注的自然语言语句拆分为疑问语义树，一个是将已标注的自然语言语句合成为疑问语义树。
- [0136] 通过统计疑问语义树及疑问语义子树，统计出相应的疑问语义句模，进而建立疑问语义句模库，其主要作用可以在实现疑问语义计算过程中，用于驱动语义规则。疑问语义句模库的作用有：
- [0137] 用来为合成句子的语义驱动库；
- [0138] 用来作为合成句子的语义驱动库；
- [0139] 用来切分和标注字符块及义项；
- [0140] 用来切分和标注未登录词；
- [0141] 用来检索句子的疑问点；
- [0142] 用来合成疑问语义树；
- [0143] 用来拆分疑问语义树；
- [0144] 用来计算自然语言语句的语义相似度；
- [0145] 例如：明天我和你一块去北京。
- [0146] 存储后为：明天/几时我和你/谁一块去/怎样北京/哪里。
- [0147] 本发明实施例通过设计多个多层次的疑问语义标注集，每个标注集均由若干现代汉语中常用的疑问词组成，其中疑问词包括疑问代词；根据待处理语句的各组成部分的疑问特征，通过所述疑问语义标注集、疑问语义句模库和疑问语义树，将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符；根据所述疑问对象或所述疑问运算符的性

质和规则,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,实现待处理语句的疑问语义计算,作为一个基础的语义计算方法,本发明实施例可以有效的解决常见的自然语言处理问题,尤其是在语句的切分标注、自然语言搜索、机器翻译、人机问答等领域具有较高的使用价值。

[0148] 实施例二

[0149] 参考图2,图2是本发明实施例基于疑问语义的自然语言语义计算的方法第二实施例的流程示意图。

[0150] 在实施例一的基础上,所述基于疑问语义的自然语言语义计算的方法还包括:

[0151] 步骤104,将搜索字符通过预先设置的算法划分为预先设置的疑问对象;

[0152] 步骤105,根据划分后的疑问代词搜索预先存储的字符;

[0153] 步骤106,若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

[0154] 具体的,例如,若用户输入几时、谁、怎样、哪儿,则根据预先存储的明天/几时我和你/谁一块去/怎样北京/哪里,可以搜索出明天我和你一块去北京。

[0155] 实施例三

[0156] 参考图3,图3是本发明实施例基于疑问语义的自然语言语义计算的方法第三实施例的流程示意图。

[0157] 在实施例一的基础上,所述方法还包括:

[0158] 步骤107,接收用户输入的搜索字符;

[0159] 步骤108,根据所述搜索字符和相似度计算获取预先存储的字符模型;

[0160] 步骤109,根据所述预先存储的字符模块划分为预先设置的疑问对象;

[0161] 步骤110,根据划分后的疑问代词搜索预先存储的字符;

[0162] 步骤111,若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。

[0163] 具体的,在处理自然语言句子时,由于对每个句子建立了相应的疑问语义树,各层子树语义间相对独立,进而实现并行计算。由于各层子树对句子的疑问语义抽象层次不同,进行计算时,可计算最低等级的字符块、也可计算抽象度递增的各层子树,进而放大或缩小了搜索空间,实现匹配精度的有效控制。

[0164] 对于具体句子的计算,可转换为与疑问语义句模库中句模型的匹配与分解问题,进而计算出语义相似度。步骤描述:

[0165] 输入句子S;

[0166] 对句子S进行疑问语义标注;

[0167] 根据句子中疑问运算符和疑问对象,进行分级计算;

[0168] 分级计算结果,根据疑问点,转化为疑问语义树;

[0169] 与疑问句模匹配,计算出各级的疑问语义句式;

[0170] 为后续处理做好准备,处理结束。

[0171] 实施例四

[0172] 参考图4,图4是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图。

[0173] 在实施例四中,所述基于疑问语义的自然语言语义计算的装置包括:

[0174] 设计模块401,用于设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;

[0175] 优选地,参考图5,图5是本发明实施例设计模块401的功能模块示意图。所述设计模块401,包括:

[0176] 设计单元501,用于根据不同的语义场景或者不同的应用场景设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词。

[0177] 标注模块402,用于根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;

[0178] 优选地,参考图6,图6是本发明实施例标注模块402的功能模块示意图。所述标注模块402,包括:

[0179] 第一标注单元601,用于若所述语义范畴为人,则与人对应的疑问代词为“谁”;

[0180] 第二标注单元602,用于若所述语义范畴为事物,则与事物对应的疑问代词为“什么”;

[0181] 第三标注单元603,用于若所述语义范畴为动作,则与所述事物对应的疑问代词为“怎么”,与所述“怎么”疑问语义等同的其他疑问代词为怎么样、怎样、怎、怎么着、怎的和如何;

[0182] 第四标注单元604,用于若所述语义范畴为时间,则与所述时间对应的疑问代词为几时,与所述“几时”疑问语义等同的其他疑问代词为多会儿、哪会儿和多咱;

[0183] 第五标注单元605,用于若所述语义范畴为地点场所,则与所述地点场所对应的疑问代词为哪里,与所述“哪里”疑问语义等同的其他疑问代词为哪和哪儿;

[0184] 第六标注单元606,用于若所述语义范畴为数目数量,则与所述数目数量对应的疑问代词为多少,与所述多少等同的其他疑问代词为几和多;

[0185] 第七标注单元607,用于若所述语义范畴为虚词,则将所述语义范畴切分并标注为疑问运算符。

[0186] 具体的,现代汉语中的疑问代词是一个相对封闭的词类。现代汉语中的疑问代词从使用频率上来看,常见的疑问代词有:什么、怎么、谁、几、哪、为什么、多少、哪儿、怎么样、何、哪里、如何、多、干吗;非常用的疑问代词:怎样、怎、怎么着、为何、怎的、几时、哪会儿、多会儿、多咱。

[0187] 鉴于以下两点:

[0188] 一个是某些方言中的疑问代词,如:啥、弄啥、干嘛、干吗、咋办、咋整等,其基本疑问语义可用现代汉语中的疑问代词来替换;另一个是组合语义的疑问代词或短语,如:什么时候、什么地方、什么人、什么事、什么东西、什么样子、什么样的、什么数目、什么数量、什么高度、什么重量、什么程度、什么情况、何人、何时、何处、何物、为何等,其基本疑问语义均能为现代汉语中的疑问代词直接或组合替代,因此不再单独罗列到疑问代词集。

[0189] 将疑问代词的基本疑问语义限定在范畴层面进行分析后,可知其能对:人、事物、时间、处所、数量、方式、性状、原因等范畴内的对象进行疑问。在疑问代词的分类上,我们采

取求同存异(动态的抓大放小)的权衡策略,即寻求其在疑问范畴层面上的一致之处(权值大),而忽略其语义或用法上的细微差异之处(权值小)。对于疑问代词,当其疑问范畴一致时,我们将其视为一类。如:“哪儿、哪里”可对表达地点类的字符块提出疑问,因此视为同类。

[0190] 对于“什么”这一特殊疑问代词,由于其可以对任何字符块提出“未知”疑问,“未知”对于了解字符块意义的贡献有限,疑问对象十分宽泛,因此我们这里采取的是缩小其疑问范畴,将其作为对表达事物类语义的字符块进行提问的疑问代词。鉴于其能跟多类字符块组合起来表达某一特定疑问范畴,1) 其组合语义可被某单一疑问代词等价替换,如:“什么地方”可被“哪里”替换;2) 涉及范畴内对象的深层次语义分析,如:“什么高度、什么宽度、什么长度”等,对于我们来说,在疑问范畴层面上属于数量范畴内的深层次语义,因此在范畴层面暂不处理;3) 要么涉及范畴内对象的综合语义分析,如:什么情况、什么原因等,此是疑问对象多个范畴层面的综合语义结果,因此在范畴层面暂不处理。

[0191] 为了利用计算机理解并处理语义,在限于范畴层面考虑疑问代词的语义时,我们从疑问语义角度入手,利用疑问代词集中的疑问代词标注疑问对象(或它的各个义项)所归属的语义范畴,即:可用哪种疑问词对疑问对象进行提问。

[0192] 根据疑问代词的疑问范畴来设计多种疑问语义标注集,以便对疑问对象进行标注,例如以下为所设计的多个多层次疑问语义标注集中具有代表性的一个:

[0193] 形式化表示为: $Y = \{谁, 什么, 怎样, 几时, 哪里, 多少, \dots\}$

[0194] 具体的,参考如下表1:

[0195]

疑问对象的语义范畴	用来标注的疑问代词	标注符号
人	谁	
事物(包含性质)	什么	
动作(包含方式)	怎么, 等价的有: 怎么样、怎样、怎、怎么着、怎的、如何	
时间	几时, 等价的有: 多会儿、哪会儿、多咱	
地点方所	哪里, 等价的有: 哪、哪儿	
数目数量	多少, 等价的有: 几、多	
无实在意义的字符块	疑问运算符	

[0196] 表1

[0197] 在自然语言处理中,语义计算是通过计算机解释自然语言中各级单位的意义,如:字、词、词组、短语、句子、句群、段落、篇章等的意义。为了处理方便,我们假定只考虑句子及组成句子的各级单位的意义。在各级单位中,假设一个有实在意义的字符块,它的意义或某个义项的意义一定属于某一或某些范畴,并且能被某一个或某些疑问代词所疑问时,称这个字符块为疑问对象。对于一些没有实在意义或无法被疑问的字符块,我们称之为疑问运算符。如:一些常见的实词为疑问对象,而一些常见的虚词为疑问运算符。每个疑问对象都作为一个疑问点,可以用于检索、人机问答以及机器翻译。

[0198] 根据疑问对象的特征或称为性质、属性将其分为若干类别,并制定若干规则。

[0199] 疑问对象的若干属性包含但不局限于:

[0200] 疑问对象或其义项所归属的疑问范畴,即用哪些疑问代词标注该疑问对象;

- [0201] 疑问对象与疑问对象的搭配属性；
- [0202] 疑问对象与疑问运算符的搭配属性；
- [0203] 疑问对象可支配的疑问对象数目(分为一元、二元、三元等)；
- [0204] 疑问对象的作用范围与运算方向；
- [0205] 同类疑问对象间、非同类疑问对象间的合并运算；
- [0206] 疑问对象间的语义侧重点；
- [0207] 疑问对象的若干运算规则(包含但不限于)：
- [0208] 分解运算；
- [0209] 复合疑问对象被分解为若干疑问对象；
- [0210] 合并运算；
- [0211] 多个疑问对象合并为一个疑问对象；
- [0212] 顺序变换运算；
- [0213] 某些疑问对象的顺序能够改变次序而保持等价语义；
- [0214] 分解运算：
- [0215] 在某些疑问对象上实施递归的疑问语义处理。
- [0216] 具体的,在句子及组成句子的各级单位中,存在非可被疑问的部分或某个义项即非疑问对象,对于这一部分可分为两类：
- [0217] 多数虚词,一般不作为疑问对象,在本发明中,称之为疑问运算符；
- [0218] 标点符号,由于其数量、用法有限,暂不考虑或特殊处理；
- [0219] 对于疑问运算符,根据其特征或称为性质分为若干类别,并根据疑问运算符的特征,制定出若干规则；
- [0220] 疑问运算符的若干属性包含但不限于：
- [0221] 疑问运算符可支配的疑问对象；
- [0222] 根据运算符可支配的疑问对象数目(分为一元、二元、三元等)；
- [0223] 运算符的作用范围；
- [0224] 运算符的运算方向从左到右,还是右到左。如“被”和“把”的一类语义句子中如考虑到方向问题,则认为:谁被什么怎么了,等价于:什么把谁怎么了。
- [0225] 疑问运算符的若干运算规则包含但不限于：
- [0226] 分解运算：
- [0227] 疑问对象被分解为疑问对象和运算符。
- [0228] 例如:疑问对象:“我和你”可以被“谁”疑问,同时作为一个复合的疑问对象,其可以分解为:疑问对象:“我”、“你”,以及疑问运算符:“和”。
- [0229] 合并运算：
- [0230] 疑问对象与运算符合并为新的疑问对象。
- [0231] 例如:同上例,疑问对象:“我”、“你”,以及疑问运算符:“和”。
- [0232] 我/谁和/疑问运算符你/谁
- [0233] 合并运算后：
- [0234] 我和你/谁
- [0235] 顺序变换运算：

- [0236] 某些疑问对象的顺序能够改变次序而保持等价语义。
- [0237] 例如:同上例,疑问对象:“我”、“你”,以及疑问运算符:“和”。
- [0238] 我和你/谁
- [0239] 顺序变换后,语义保持不变:
- [0240] 你和我/谁
- [0241] 递归运算:
- [0242] 在某些疑问对象上实施递归的疑问语义处理。
- [0243] 计算模块403,用于根据所述疑问对象或所述疑问运算符的性质和规则,结合所述疑问语义句模库、统计方法和疑问语义树,实现待处理语句的疑问语义计算。
- [0244] 具体的,在对句子的字符块进行标注后,我们利用疑问对象性质、运算符性质以及运算规则对其进行处理,进而建立起疑问语义树。将句子中的疑问对象作为疑问点,利用疑问点可以用来回答相应的搜索、翻译或人机对话。疑问点对应疑问语义树中的节点。对于疑问语义树,我们用其节点表示字符块,用边表示标注符号。在这里,我们可以通过分解与合并来操作疑问语义树,一个是将未标注的自然语言语句拆分为疑问语义树,一个是将已标注的自然语言语句合成为疑问语义树。
- [0245] 通过统计疑问语义树及疑问语义子树,统计出相应的疑问语义句模,进而建立疑问语义句模库,其主要作用可以在实现疑问语义计算过程中,用于驱动语义规则。疑问语义句模库的作用有:
- [0246] 用来为合成句子的语义驱动库;
- [0247] 用来作为合成句子的语义驱动库;
- [0248] 用来切分和标注字符块及义项;
- [0249] 用来切分和标注未登录词;
- [0250] 用来检索句子的疑问点;
- [0251] 用来合成疑问语义树;
- [0252] 用来拆分疑问语义树;
- [0253] 用来计算自然语言语句的语义相似度;
- [0254] 例如:明天我和你一块去北京。
- [0255] 存储后为:明天/几时我和你/谁一块去/怎样北京/哪里。
- [0256] 本发明实施例通过设计多个多层次的疑问语义标注集,每个标注集均由若干现代汉语中常用的疑问词组成,其中疑问词包括疑问代词;根据待处理语句的各组成部分的疑问特征,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,将所述待处理语句的各组成部分切分并标注为疑问对象或疑问运算符;根据所述疑问对象或所述疑问运算符的性质和规则,通过所述疑问语义标注集、疑问语义句模库和疑问语义树,实现待处理语句的疑问语义计算,作为一个基础的语义计算方法,本发明实施例可以有效的解决常见的自然语言处理问题,尤其是在语句的切分标注、自然语言搜索、机器翻译、人机问答等领域具有较高的使用价值。
- [0257] 实施例五
- [0258] 参考图7,图7是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图。

- [0259] 在实施例三的基础上,所述装置还包括:
- [0260] 第一划分模块404,用于将搜索字符通过预先设置的算法划分为预先设置的疑问对象;
- [0261] 第一搜索模块405,用于根据划分后的疑问代词搜索预先存储的字符;
- [0262] 第一显示模块406,用于若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。
- [0263] 具体的,例如,若用户输入几时、谁、怎样、哪儿,则根据预先存储的明天/几时我和你/谁一块去/怎样北京/哪里,可以搜索出明天我和你一块去北京。
- [0264] 实施例六
- [0265] 参考图8,图8是本发明实施例基于疑问语义的自然语言语义计算的装置的功能模块示意图。
- [0266] 在实施例四的基础上中,所述装置还包括:
- [0267] 接收模块407,用于接收用户输入的搜索字符;
- [0268] 获取模块408,用于根据所述搜索字符和相似度计算获取预先存储的字符模型;
- [0269] 第二划分模块409,用于根据所述预先存储的字符模块划分为预先设置的疑问对象;
- [0270] 第二搜索模块410,用于根据划分后的疑问代词搜索预先存储的字符;
- [0271] 第二显示模块411,用于若划分后的疑问代词和预先存储的字符对应,则显示与所述预先存储的字符对应的划分前的待处理的字符。
- [0272] 具体的,在处理自然语言句子时,由于对每个句子建立了相应的疑问语义树,各层子树语义间相对独立,进而实现并行计算。由于各层子树对句子的疑问语义抽象层次不同,进行计算时,可计算最低等级的字符块、也可计算抽象度递增的各层子树,进而放大或缩小了搜索空间,实现匹配精度的有效控制。
- [0273] 对于具体句子的计算,可转换为与疑问语义句模库中句模型的匹配与分解问题,进而计算出语义相似度。步骤描述:
- [0274] 输入句子S;
- [0275] 对句子S进行疑问语义标注;
- [0276] 根据句子中疑问运算符和疑问对象,进行分级计算;
- [0277] 分级计算结果,根据疑问点,转化为疑问语义树;
- [0278] 与疑问句模匹配,计算出各级的疑问语义句式;
- [0279] 为后续处理做好准备,处理结束。
- [0280] 以上结合具体实施例描述了本发明实施例的技术原理。这些描述只是为了解释本发明实施例的原理,而不能以任何方式解释为对本发明实施例保护范围的限制。基于此处的解释,本领域的技术人员不需要付出创造性的劳动即可联想到本发明实施例的其它具体实施方式,这些方式都将落入本发明实施例的保护范围之内。

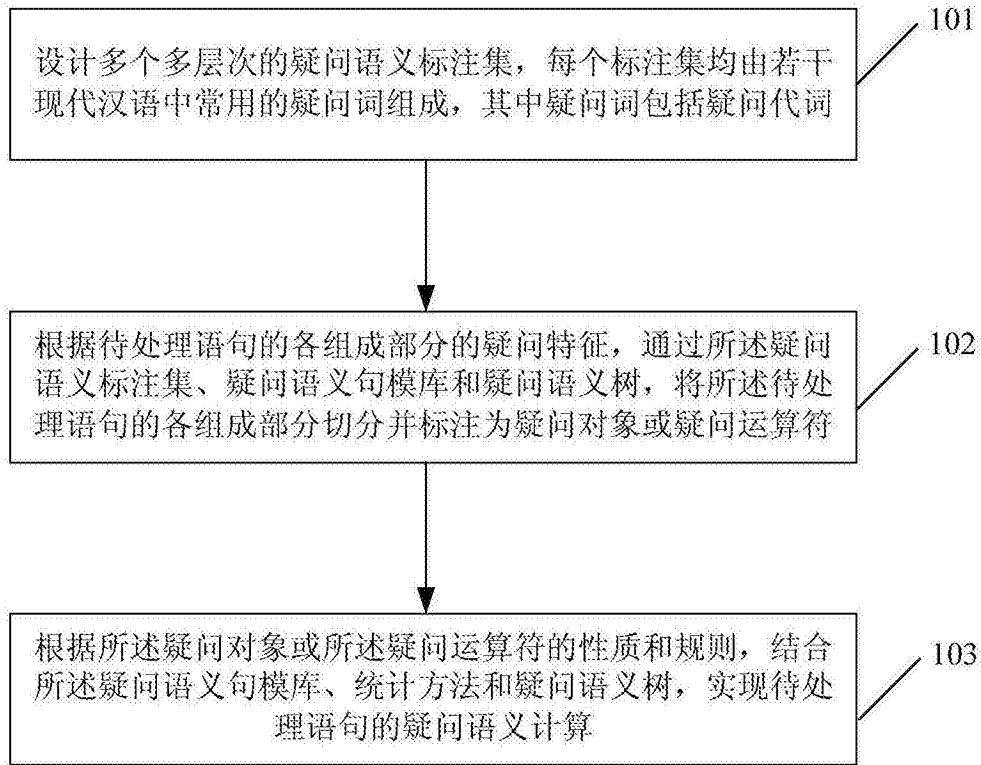


图1

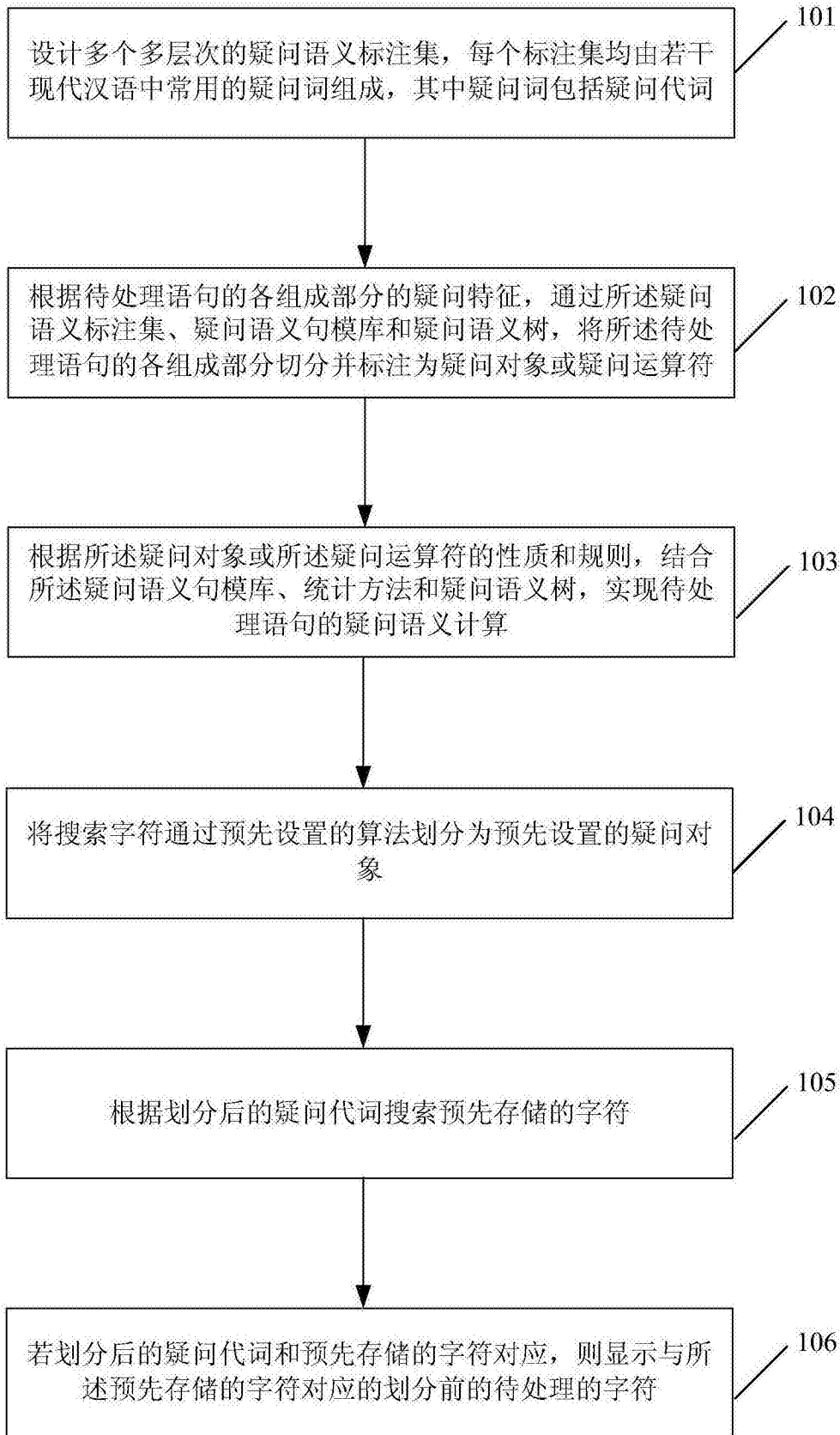


图2

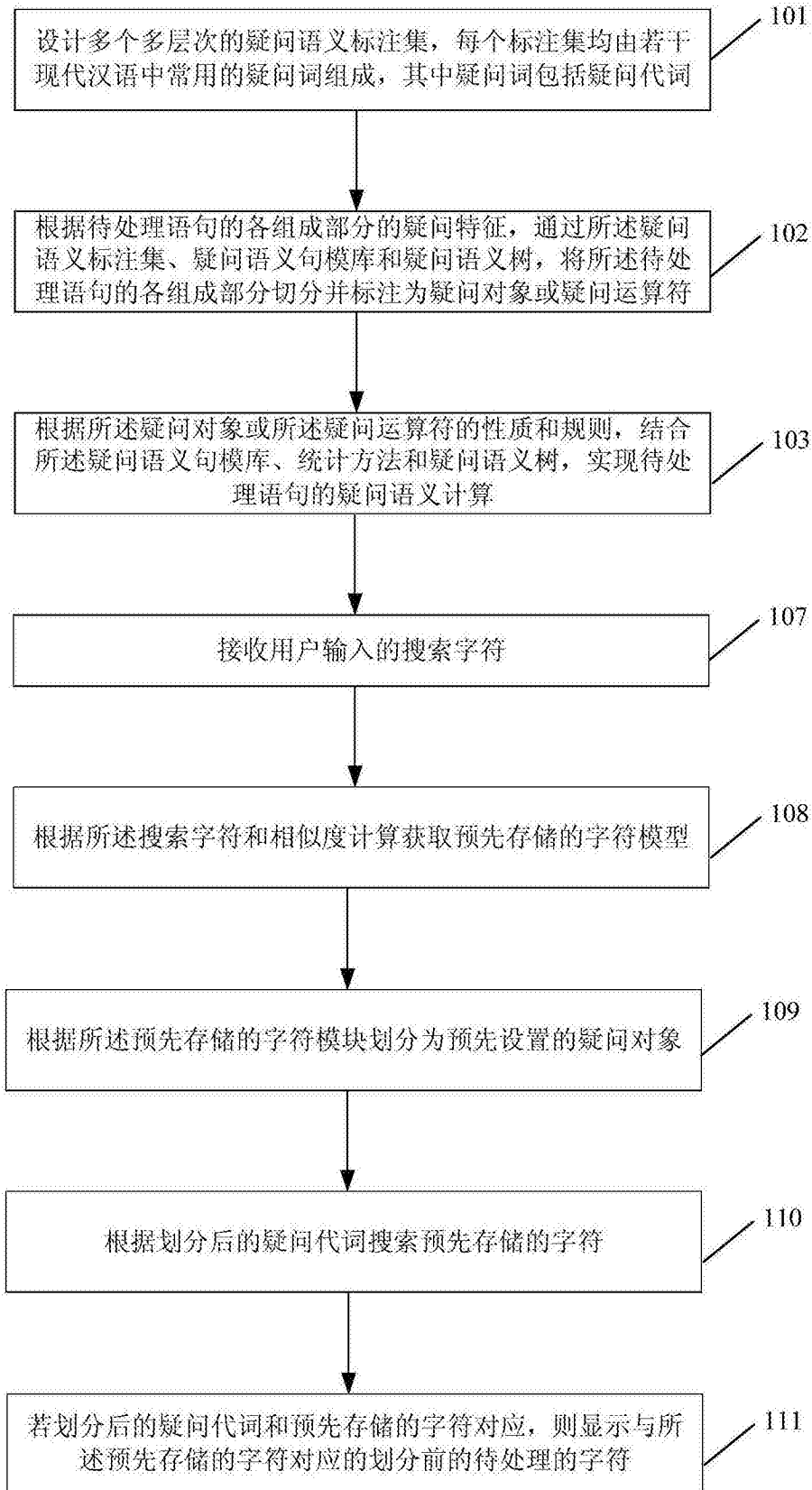


图3

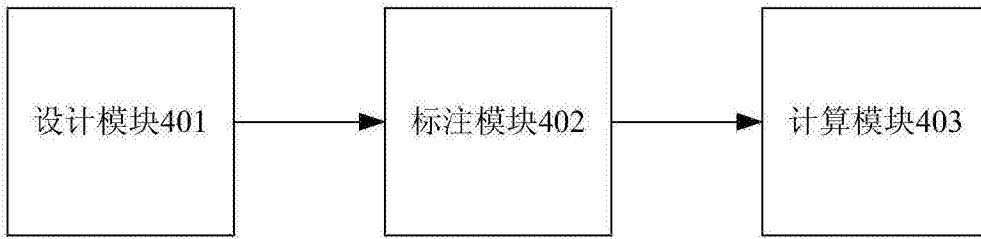


图4

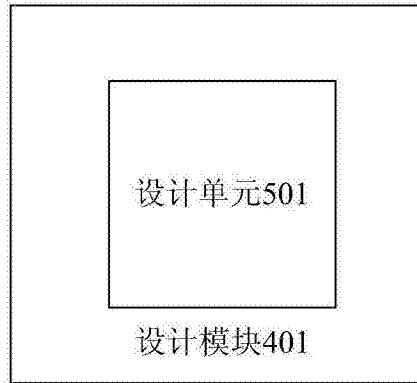


图5

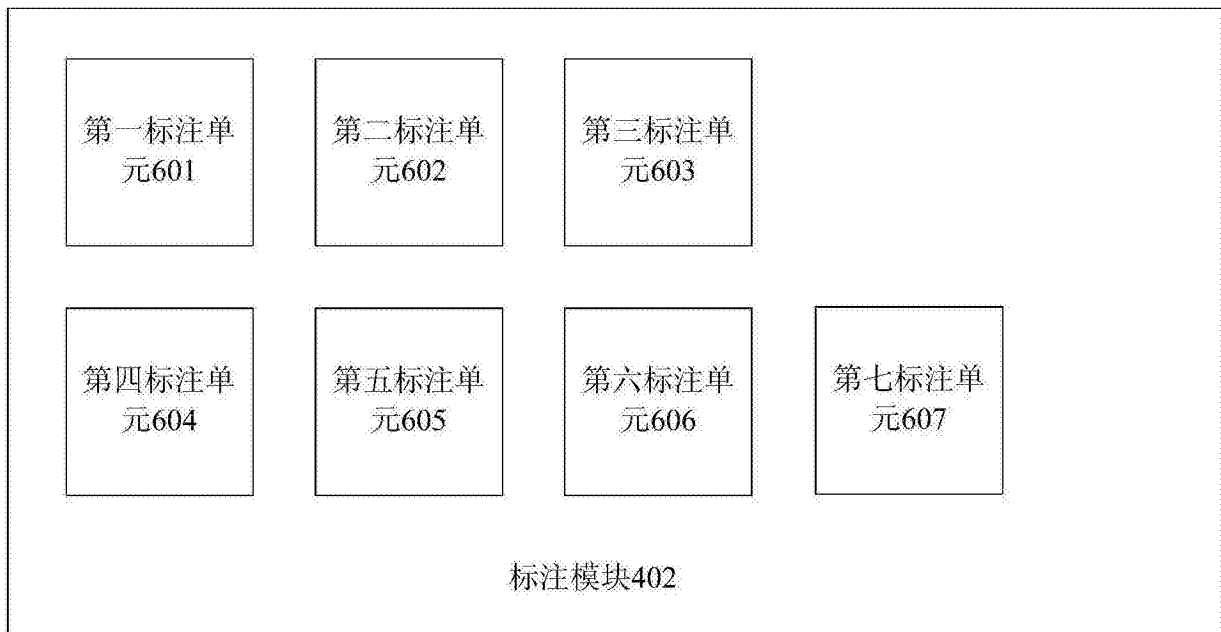


图6

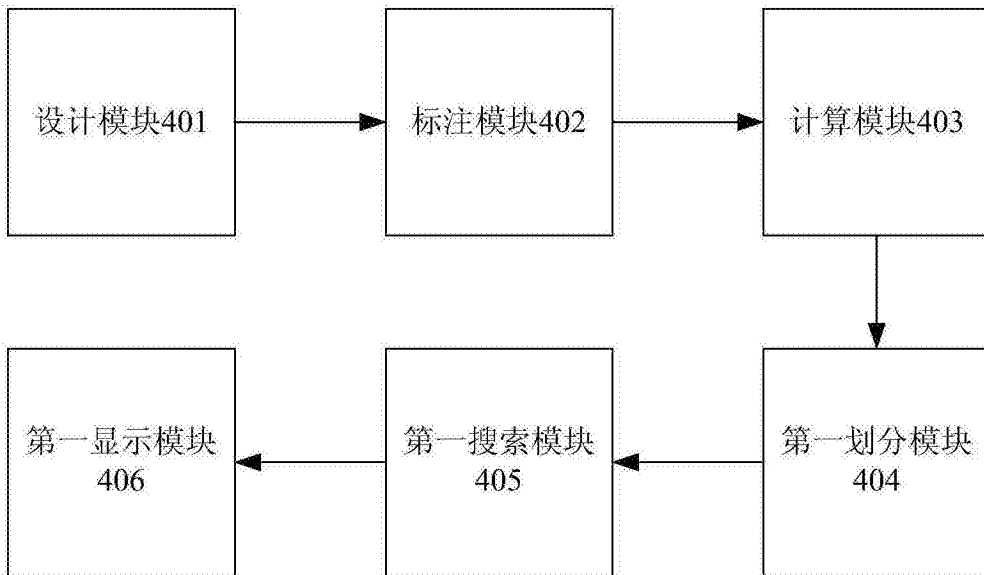


图7

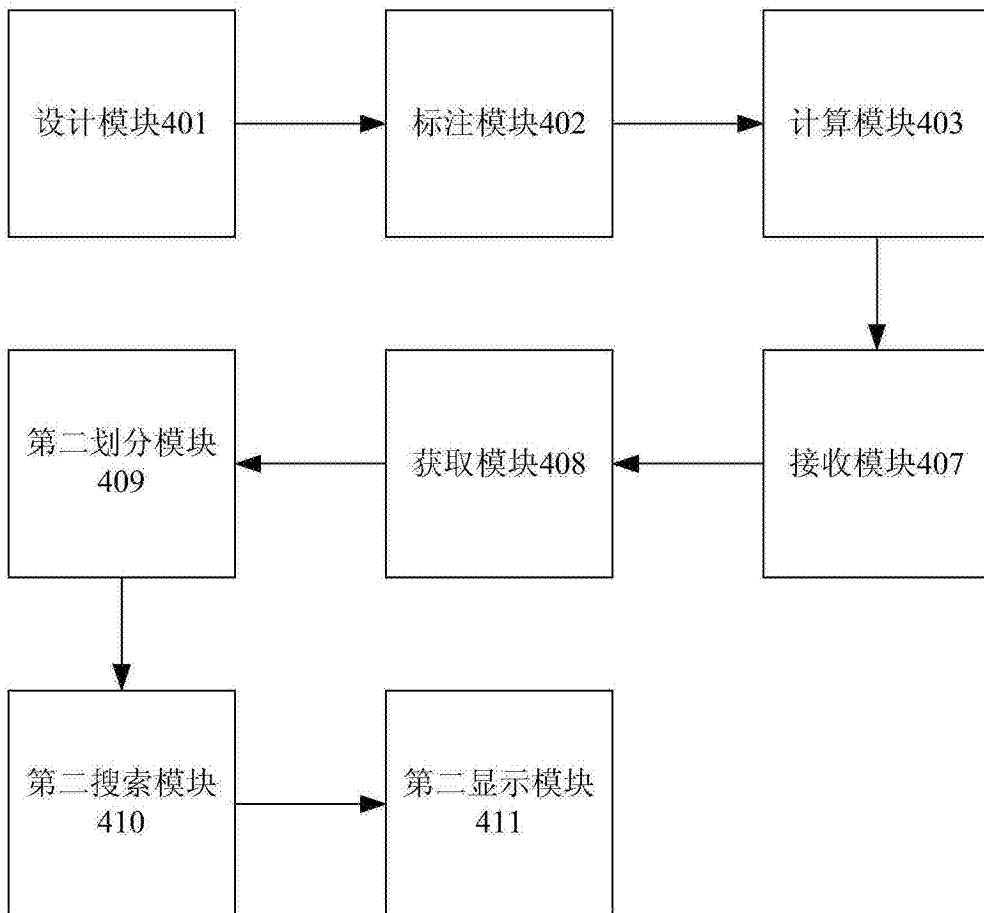


图8