(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104978396 A
(43)申请公布日 2015.10.14

(21)申请号 201510295701.7

(22)申请日 2015.06.02

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 王波 田力 李羽

(74)专利代理机构 广州三环专利代理有限公司 44202
代理人 温旭 郝传鑫

(51)Int.Cl.
G06F 17/30(2006.01)
G09B 7/02(2006.01)

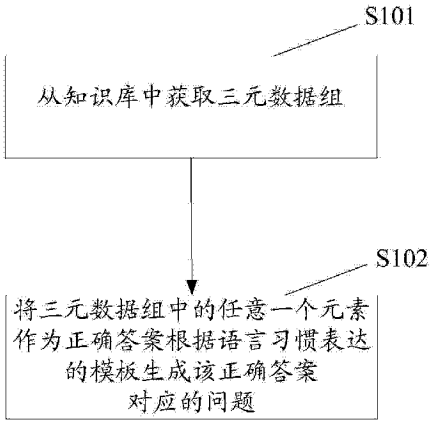
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种基于知识库的问答题目生成方法和装置

(57)摘要

本发明提供了一种基于知识库的问答题目生成方法和装置。该方法可包括：从知识库中获取三元数据组，其中，所述三元数据组包括实体、属性和属性值三个元素，各元素的关系为：属性值元素为实体元素的属性元素对应的取值；将该三元数据组中的任意一个元素作为正确答案根据语言习惯表达的模板生成该正确答案对应的问题。本发明的上述方法和装置，能够以三元数据组中的一个元素作为正确答案，另外两个元素转化为对应问题的主干，根据语言习惯表达模板显著改善了海量的结构化知识数据向问答题目的数据转化效率。



CN 104978396 A

1. 一种基于知识库的问答题目生成方法，其特征在于，包括：

从知识库中获取三元数据组，其中，所述三元数据组包括实体、属性和属性值三个元素，各元素的关系为：属性值元素为实体元素的属性元素对应的取值；

将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题。

2. 根据权利要求 1 所述的方法，其特征在于，将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题，包括：

根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成。

3. 根据权利要求 2 所述的方法，其特征在于，将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题，包括：

在根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成之后，

将所述三元数据组中选择出的属性对应的属性值元素作为正确答案，

根据语言习惯表达的模板生成包含下述句子成分的问句作为所述正确答案对应的问题：所述三元数据组中实体的属性。

4. 根据权利要求 1 至 3 中任意一项所述的方法，其特征在于，还包括：

根据所述正确答案和 / 或生成的所述问题的约束信息生成所述问题的一个以上错误答案，所述约束信息包括下述的一种以上：属性集合、属性值集合、热门程度、公知常识。

5. 根据权利要求 4 所述的方法，其特征在于，还包括：

对所述正确答案和 / 或所述错误答案进行配图。

6. 根据权利要求 5 所述的方法，其特征在于，还包括：

在所述问题的显示页面呈现正确答案查看链接，供用户查看所述问题的正确答案。

7. 一种基于知识库的问答题目生成装置，其特征在于，包括：

获取模块，用于从知识库中获取三元数据组，其中，所述三元数据组包括实体、属性和属性值三个元素，各元素的关系为：属性值元素为实体元素的属性元素对应的取值；

生成模块，用于将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题。

8. 根据权利要求 7 所述的装置，其特征在于，所述生成模块中将所述三元数据组中的任意一个元素作为正确答案根据语言习惯表达的模板生成所述正确答案对应的问题包括：根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成。

9. 根据权利要求 8 所述的装置，其特征在于，所述生成模块中将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题包括：在根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成之后，将所述三元数据组中选择出的属性对应的属性值元素作为正确答案，根据语言习惯表达的模板生成包含下述句子成分的问句作为所述正确答案对应的问题：所述三元数据组中实体的属性。

10. 根据权利要求 7 至 9 中任意一项所述的装置，其特征在于，还包括：

误选模块，用于根据所述正确答案和 / 或生成的所述问题的约束信息生成所述问题的一个以上错误答案，所述约束信息包括下述的一种以上：属性集合、属性值集合、热门程度、公知常识。

11. 根据权利要求 10 所述的装置，其特征在于，还包括：

配图模块，用于对所述正确答案和 / 或所述错误答案进行配图。

12. 根据权利要求 11 所述的装置，其特征在于，还包括 ：

查看模块，用于在所述问题的显示页面呈现正确答案查看链接，供用户查看所述问题的正确答案。

一种基于知识库的问答题目生成方法和装置

技术领域

[0001] 本发明涉及互联网领域，具体而言，涉及一种基于知识库的问答题目生成方法和装置。

背景技术

[0002] 通过问题和对该问题的回答是人们获取现实世界的知识数据或信息的一种有效方式。然而，现实世界（特别是互联网上）涉及的知识数据或信息是海量的，如何将海量的知识数据或信息转换为问题和相应回答成为亟需解决的技术问题。现有的知识数据或信息到问题和相应回答的转换主要通过人工方式，转换的数据处理效率较低。

发明内容

[0003] 为解决上述的技术问题，本发明提供了一种基于知识库的问答题目生成方法和装置，利用海量的结构化的实体、属性以及属性值信息作为问题的主干数据，将实体、属性以及属性值三元素之一作为正确答案，剩余两个元素作为问题的问句成分，与现有的人工编辑生成问答题库相比，显著提高了海量的知识数据向问答题目的数据转换效率并且改善了问答题目的丰富度；而且，通过海量结构化数据自动生成问答题库，能够避免人工编辑问题的记忆偏差，提高答案的准确度。

[0004] 根据本发明实施方式的第一方面，提供了一种基于知识库的问答题目生成方法，该方法可包括：从知识库中获取三元数据组，其中，所述三元数据组中各元素的关系为：属性值元素为实体元素的属性元素对应的取值；将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题。

[0005] 在本发明的一些实施方式中，将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题，包括：根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成。

[0006] 在本发明的一些实施方式中，将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题，包括：在根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成之后，将所述三元数据组中选择出的属性对应的属性值元素作为正确答案，根据语言习惯表达的模板生成包含下述句子成分的问句作为所述正确答案对应的问题：所述三元数据组中实体的属性。

[0007] 在本发明的一些实施方式中，基于知识库的问答题目生成方法还可包括：根据所述正确答案和／或生成的所述问题的约束信息生成所述问题的一个以上错误答案，所述约束信息包括下述的一种以上：属性集合、属性值集合、热门程度、公知常识。

[0008] 在本发明的一些实施方式中，基于知识库的问答题目生成方法还可包括：对所述正确答案和／或所述错误答案进行配图。

[0009] 在本发明的一些实施方式中，基于知识库的问答题目生成方法还可包括：在所述问题的显示页面呈现正确答案查看链接，供用户查看所述问题的正确答案。

[0010] 根据本发明实施方式的第二方面，提供了一种基于知识库的问答题目生成装置，该装置可包括：获取模块，用于从知识库中获取三元数据组，其中，所述三元数据组中各元素的关系为：属性值元素为实体元素的属性元素对应的取值；生成模块，用于将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题。

[0011] 在本发明的一些实施方式中，所述生成模块中将所述三元数据组中的任意一个元素作为正确答案根据语言习惯表达的模板生成所述正确答案对应的问题包括：根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成。

[0012] 在本发明的一些实施方式中，所述生成模块中将所述三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成所述正确答案对应的问题包括：在根据所述三元数据组中属性的点击数量选择一个以上属性用于问题生成之后，将所述三元数据组中选择出的属性对应的属性值元素作为正确答案，根据语言习惯表达的模板生成包含下述句子成分的问句作为所述正确答案对应的问题：所述三元数据组中实体的属性。

[0013] 在本发明的一些实施方式中，基于知识库的问答题目生成装置还可包括：误选模块，用于根据所述正确答案和/或生成的所述问题的约束信息生成所述问题的一个以上错误答案，所述约束信息包括下述的一种以上：属性集合、属性值集合、热门程度和公知常识。

[0014] 在本发明的一些实施方式中，基于知识库的问答题目生成装置还可包括：配图模块，用于对所述正确答案和/或所述错误答案进行配图。

[0015] 在本发明的一些实施方式中，基于知识库的问答题目生成装置还可包括：查看模块，用于在所述问题的显示页面呈现正确答案查看链接，供用户查看所述问题的正确答案。

[0016] 本发明实施方式提供的知识数据转换方法和装置，利用实体、属性、属性值等结构化数据生成相应的问题和答案，与现有的问题和答案的人工生成技术相比，改善了知识数据的数据转换效率，同时也提高了问答题目的丰富度；通过问题和答案的约束信息生成问题的错误选项，增加了问答题目的层次性，有利于促使用户参与答题，获取知识；对答案选项进行相应的配图，改善了问答题目的趣味性，更进一步改善用户参与答题的积极性；通过在问题页面设置正确答案查看链接，有利于用户实时查看答案信息，获取相关知识。

附图说明

[0017] 图 1 图示了根据本发明一种实施方式的基于知识库的问答题目生成方法的流程示意图；

[0018] 图 2 图示了根据本发明一种实施方式的问题和答案页面的网页示意图；

[0019] 图 3 图示了根据本发明一种实施方式的基于知识库的问答题目生成装置的结构示意图。

具体实施方式

[0020] 为使本发明实施方式的目的、技术方案和优点更加清楚，下面将结合附图对本发明作进一步地详细描述。

[0021] 参见图 1，图示了根据本发明一种实施方式的基于知识库的问答题目生成方法的流程示意图，该方法可包括：

[0022] S101，从知识库中获取三元数据组，其中，所述三元数据组包括实体、属性和属性值三个元素。各元素的关系为：属性值元素为实体元素的属性元素对应的取值；

[0023] S102，将三元数据组中的任意一个元素作为正确答案根据语言表达习惯的模板生成该正确答案对应的问题。

[0024] 基于知识库的问答题目生成方法可包括执行步骤S101，从知识库中获取三元数据组＜实体，属性，属性值＞，其中，所述三元数据组中各元素的关系为：属性值元素为实体元素的属性元素对应的取值。具体而言，可包括：根据属性/实体词典获取二元数据组＜实体，属性＞；根据二元数据组＜实体，属性＞从知识库中获取二元数据组中实体的属性对应的属性值，将该属性值与该二元数据组合并生成三元数据组＜实体，属性，属性值＞。根据属性/实体词典获取二元数据组＜实体，属性＞，从知识库中获取属性/实体词典，其中，属性/实体的选择可通过圈定或选定一些特定的领域和实体进行，例如，首先，可根据领域的属性热度、数据热度、实体热度、用户主需求（其中，这些热度可依据用户的检索、点击量等因素进行测算）等计算用户对某一些或某一个领域的兴趣度，圈出这些或这个领域作为问答题目的生成领域；然后，根据圈出的这些领域中的实体热度圈出该领域中的实体，作为知识数据转换的实体。利用上述圈定的领域及其实体生成问题，可提高问题的兴趣度，促进用户参与答题，获取相关知识。根据上述圈定的领域和实体，依据属性/实体词典，形成二元组数据＜实体，属性＞。为了方便描述起见，本文用符号S表示实体，符号P表示属性。实体和属性在通常的语言表达中以“实体的属性”形式成对出现，例如，实体为男明星A，他相应的属性可包括：性别、民族、年龄、妻子等，在通常的语言中表达为：实体的属性，例如，男明星A的性别、男明星A的民族、男明星A的年龄、男明星A的妻子等。

[0025] 根据上述获取的二元数据组＜实体，属性＞，可以通过机器学习之类的算法进行训练，挖掘语言表达习惯的模板。利用挖掘出的语言表达的模板，可以对二元数据组中的实体和属性元素进行问句改写，生成对应的问题。

[0026] 根据获取的二元数据组＜实体，属性＞从知识库中获取该二元数据组中实体的属性对应的属性值，并将该属性值与二元数组合并生成三元数据组＜实体，属性，属性值＞。从大规模的知识库（例如，语料）中可获取二元数据组＜实体，属性＞中实体的属性对应的属性值，例如，对于二元数据组＜男明星A，妻子＞，从大规模的语料中可获得该二元数据组中男明星A（实体）的妻子（属性）对应的属性值：女明星B。然后，将获取的属性值（女明星B）与二元数据组＜男明星A，妻子＞合并得到三元数据组＜男明星A，妻子，女明星B＞。

[0027] 接下来，执行步骤S102，将步骤S101获取的三元数据组＜实体，属性，属性值＞中的任意一个元素作为正确答案根据语言表达习惯的模板生成该正确答案对应的问题。例如，可以将三元数据组＜实体，属性，属性值＞中的实体、属性、属性值三者中的任意一者作为正确答案，另外二者作为问题的主干结构，生成该正确答案对应的问题，例如，可以生成多种问句形式的问题。

[0028] 在一些实施方式中，可根据三元数据组＜实体，属性，属性值＞中的实体和属性，获取查询日志(querylog)，对querylog中涉及的实体和属性进行过滤（例如，过滤的逻辑可以为：querylog需要包含一个实体和一个属性），然后按照属性点击量进行累加，得到属性的重要度数据。例如，利用查询日志的点击量的累加统计，也就是搜索次数的加权和，累加是按照相同的属性累计。对于：Query1：点击加权5.4包含实体S1和P1；Query1：点击

加权 1.4 包含实体 S1 和 P2 ;Query1: 点击加权 3.4 包含实体 S1 和 P1。那么 P1 的累加是 5.4+3.4，P2 的累加是 1.4。属性累加的值越高说明用户对该属性越关注，该属性越重要。利用这样的重要度较高的属性生成问题，对用户而言，兴趣度较高，也可认为生成问题的质量较高。在获取属性重要度数据的情形下，可以利用这些属性生成问句成分包括“实体的属性”的问题。例如，实体（男明星 A）的属性之一妻子的重要度比较高，那么可以生成下述这样的问题：“男明星 A 的妻子是？”、“谁是男明星 A 的妻子？”等。

[0029] 在一些实施方式中，还可根据二元数据组＜实体，属性＞挖掘人们的语言表达习惯，然后利用三元数据组＜实体，属性，属性值＞中的一个元素作为正确答案生成问题。例如，可以根据挖掘的语言表达习惯将实体、属性、属性值中的两者作为主干结构生成问句，也就是说，依据语言表达习惯的模板对三元数据中的元素进行问句改写。例如，对于三元数据组＜男子 A，儿子，男孩 B>，可以改写为下述形式的问题：“男子 A 的儿子是？”、“谁是男子 A 的儿子？”、“男子 A 的儿子名字是？”、“谁的儿子是男孩 B”等。

[0030] 在一些实施方式中，还可以根据属性值的不同类型（例如，属性值为人物类型，或者属性值为列表），生成另外形式的问句。在属性值为人物类型的情况下，以该属性值作为正确答案生成的问题可以是：“哪位是实体的属性？”。例如，对于三元数据组＜男明星 A，妻子，女明星 B>，属性值为女明星 B 的情况下，以该属性值作为正确答案生成的问题可以是：“哪位是男明星 A 的妻子？”。在属性值为列表的情况下，以该属性值作为正确答案生成的问题可以是：“哪个是实体的属性？”或者“哪个不是实体的属性？”。例如，哪个是 XX 的好友？或者，哪个不是 XX 的好友？。

[0031] 在本发明的一些实施方式，基于知识库的问答题目生成方法还可以包括：根据正确答案和 / 或生成的问题的约束信息生成该问题的一个以上错误答案，其中，约束信息可包括正确答案的约束信息、生成的问题的约束信息以及正确答案和生成的问题的共同的约束信息。约束信息的种类可包括，但不限于下述的一种或多种（例如，大于或等于 2 种）：属性集合、属性值结合、热门程度以及公知常识。

[0032] 例如，属性值元素作为正确答案，问题为“实体的属性是？”的情形下，可以利用属性值集合约束信息、属性集合约束信息或属性值集合和属性集合约束信息三种中的一种、二种或三种。对于属性值集合约束信息，可以通过查找生成该问题的三元数据组中的属性值所对应的属性值的集合和生成该问题的三元数据组中的实体的属性对应的属性值的集合的差集，利用共现分析辅助生成。也就是说，从上述属性值的差集中选取与作为正确答案的属性值相似的选项作为错误选项，可选择多个这样的错误选项。例如，具有下述三元数据组：＜男明星 A、妻子、女明星 a>＜男明星 B、妻子、女明星 b>，＜男明星 C、妻子、女明星 c>，＜男明星 D、妻子、女明星 d>；对于问题：男明星 A 的妻子？，生成该问题的三元数据组＜男明星 A、妻子、女明星 a> 中的属性值对应的集合为 S1，即｛女明星 a｝，生成该问题的实体的属性对应的属性值的集合为 S2，即｛女明星 a，女明星 b，女明星 c，女明星 d｝。集合 S1 和 S2 的差集为｛女明星 b，女明星 c，女明星 d｝。利用这样的差集可生成相似错误选项：女明星 b，女明星 c，女明星 d。

[0033] 对于属性集合约束信息，可以利用夫妻属性、出生日期属性等作为约束信息。例如，在属性为妻子或丈夫的情况下，可以限定错误选项为女性或男性。又例如，在属性为出生日期的情况下，可以限定出生日期大于 1900 等。

[0034] 对于热门程度（例如，可通过用户搜索量或点击量等因子进行量化）约束信息，可以用于对上面获取的相似的错误选项进行筛选，筛选掉热门程度较低的错误选项。在一些实施方式中，可以省略上面的相似选项选取步骤，仅利用热门程度生成热门程度较高的错误选项。

[0035] 对于公知常识约束信息，例如，在实体为“XX 妹妹”的情况下，她的性别属性，根据公知常识可知为女性，她的年龄不会是 40 岁以上。

[0036] 在本发明的实施方式中，基于知识库的问答题目生成方法还可包括：对正确答案和 / 或错误答案进行配图。通过上述的描述可知，正确答案或错误答案可以是实体、属性或属性值中的一种，对正确答案和 / 或错误的答案的配图可包括对实体、属性或属性值的配图。对于实体的配图，例如，对于人物实体，可以采用该人物的头像、剧照等图片。对于属性值（例如，星座、国籍、属相、职业、民族等）的配图，例如，对于国籍的配图，可以采用该国的国旗等图片。

[0037] 在本发明的实施方式中，基于知识库的问答题目生成方法还可包括：在生成的问题的显示页面呈现正确答案查看链接，供用户查看该问题的正确答案。例如，可以如图 2 所示，在问题的后面设置“偷看答案”选项，作为正确答案的查看链接，点击该连接，用户可查看该问题的正确答案，方便了用户实时了解自己的答题状况。

[0038] 通过上面的描述，通过海量知识库的结构化数据可生成问题、正确答案和错误答案，可与不同的类别关联，而且同一题目可属于多个类别。例如，该问题可以与正确答案关联，也可以与题目的实体或属性关联。

[0039] 本发明的实施方式还可以包括对生成的问题进行评估，例如，根据实体热度、问题分类、点击需求、搜索结果分布、UGC(User Generated Content，用户生成内容）等兴趣建立随机森林机器学习模板预计实体的兴趣度；并根据问题热度、问题惊喜度、问题时效性建立逻辑回归机器学习模板预测问题的兴趣度。同时，还可利用线上的点击日志分析实体和题目的兴趣度。

[0040] 以上结合具体实施方式描述了本发明的基于知识库的问答题目生成方法的流程，下面将结合具体实施方式描述应用上述转换方法的基于知识库的问答题目生成装置。

[0041] 参见图 3，图示了根据本发明一种实施方式的基于知识库的问答题目生成装置的结构示意图，该装置 200 可包括：

[0042] 获取模块 201，用于用于从知识库中获取三元数据组，其中，所述三元数据组包括实体、属性和属性值三个元素，各元素的关系为：属性值元素为实体元素的属性元素对应的取值；

[0043] 生成模块 202，用于将该三元数据组中的任意一个元素作为正确答案生成该正确答案对应的问题。

[0044] 本发明实施方式的基于知识库的问答题目生成装置 200 可包括获取模块 201 和生成模块 202，可设置于物理上分离的多个位置。这些模块可与存储知识数据的知识库通讯连接，获取知识库中的相关知识数据。

[0045] 获取模块 201 可从知识库中获取三元数据组 < 实体，属性，属性值 >，其中，所述三元数据组中各元素的关系为：属性值元素为实体元素的属性元素对应的取值。具体而言，可包括：根据属性 / 实体词典获取二元数据组 < 实体，属性 >；根据二元数据组 < 实体，属性 >

从知识库中获取二元数据组中实体的属性对应的属性值，将该属性值与该二元数据组合并生成三元数据组＜实体，属性，属性值＞。可从知识库中获取属性／实体词典，其中，属性／实体的选择可通过圈定领域和实体进行，例如，首先，可根据领域的属性热度、数据热度、实体热度、用户主需求（其中，这些热度可依据用户的检索、点击量等因素进行测算）等计算用户对某一些领域的兴趣度，圈出这些领域作为问答题目生成领域；然后，根据圈出的这些领域中的实体热度圈出该领域中的实体，作为知识数据转换的实体。利用上述圈定的领域及其实体生成问题，可提高问题的兴趣度，促进用户参与答题，获取相关知识。根据上述圈定的领域和实体，依据属性／实体词典，形成二元组数据＜实体，属性＞。

[0046] 根据上述获取的二元数据组＜实体，属性＞，可以通过机器学习之类的算法进行训练，挖掘语言表达习惯的模板。利用挖掘出的语言表达的模板，可以对二元数据组中的实体和属性元素进行问句改写，生成对应的问题。

[0047] 根据获取的二元数据组＜实体，属性＞从知识库中获取该二元数据组中实体的属性对应的属性值，并将该属性值与二元数组合并生成三元数据组＜实体，属性，属性值＞。从大规模的知识库（例如，语料）中可获取二元数据组＜实体，属性＞中实体的属性对应的属性值，例如，对于二元数据组＜男明星 A，妻子＞，从大规模的语料中可获得该二元数据组中男明星 A（实体）的妻子（属性）对应的属性值：女明星 B。然后，将获取的属性值（女明星 B）与二元数据组＜男明星 A，妻子＞合并得到三元数据组＜男明星 A，妻子，女明星 B＞。

[0048] 生成模块 202 获取的三元数据组＜实体，属性，属性值＞中的任意一个元素作为正确答案根据语言表达习惯的模板生成该正确答案对应的问题。例如，可以将三元数据组＜实体，属性，属性值＞中的实体、属性、属性值三者中的任意一者作为正确答案，另外二者作为问题的主干结构，生成该正确答案对应的问题。对于三元数据组＜实体，属性，属性值＞，将属性值元素作为正确答案，将实体元素和属性元素作为句子主干成分：实体的属性，将句子的主干成分构成该正确答案的问题，例如，例如，可以生成多种问句形式的问题。

[0049] 在一些实施方式中，可根据三元数据组＜实体，属性，属性值＞中的实体和属性，获取查询日志（querylog），对 querylog 中涉及的实体和属性进行过滤（例如，过滤的逻辑可以为：querylog 需要包含一个实体和一个属性），然后按照属性点击量进行累加，得到属性的重要度数据。例如，利用查询日志的点击量的累加统计，也就是搜索次数的加权和，累加是按照相同的属性累计。对于：Query1：点击加权 5.4 包含实体 S1 和 P1；Query1：点击加权 1.4 包含实体 S1 和 P2；Query1：点击加权 3.4 包含实体 S1 和 P1。那么 P1 的累加是 5.4+3.4，P2 的累加是 1.4。属性累加的值越高说明用户对该属性越关注，该属性越重要。利用这样的重要度较高的属性生成问题，对用户而言，兴趣度较高，也可认为生成问题的质量较高。在获取属性重要度数据的情形下，可以利用这些属性生成问句成分包括“实体的属性”的问题。例如，实体（男明星 A）的属性之一妻子的重要度比较高，那么可以生成下述这样的问题：“男明星 A 的妻子是？”、“谁是男明星 A 的妻子？”等。

[0050] 在一些实施方式中，还可根据二元数据组＜实体，属性＞挖掘人们的语言表达习惯，然后利用三元数据组＜实体，属性，属性值＞中的一个元素作为正确答案生成问题。例如，可以根据挖掘的语言表达习惯将实体、属性、属性值中的两者作为主干结构生成问句，也就是说，依据语言表达习惯表达习惯的模板对三元数据中的元素进行问句改写。例如，对于三元数据组＜男子 A，儿子，男孩 B＞，可以改写为下述形式的问题：“男子 A 的儿子是？”、

“谁是男子 A 的儿子？”、“男子 A 的儿子名字是？”、“谁的儿子是男孩 B”等。

[0051] 在一些实施方式中，还可以根据属性值的不同类型（例如，属性值为人物类型，或者属性值为列表），生成另外形式的问句。在属性值为人物类型的情况下，以该属性值作为正确答案生成的问题可以是：“哪位是实体的属性？”。例如，对于三元数据组＜男明星 A，妻子，女明星 B＞，属性值为女明星 B 的情况下，以该属性值作为正确答案生成的问题可以是：“哪位是男明星 A 的妻子？”。在属性值为列表的情况下，以该属性值作为正确答案生成的问题可以是：“哪个是实体的属性？”或者“哪个不是实体的属性？”。例如，哪个是 XX 的好友？或者，哪个不是 XX 的好友？。

[0052] 在本发明的一些实施方式，基于知识库的问答题目生成装置 200 还可以包括误选模块，该误选模块根据正确答案和 / 或生成的问题的约束信息生成该问题的一个以上错误答案，其中，约束信息可包括正确答案的约束信息、生成的问题的约束信息以及正确答案和生成的问题的共同的约束信息。约束信息的种类可包括，但不限于下述的一种或多种（例如，大于或等于 2 种）：属性集合、属性值结合、热门程度和公知常识。例如，属性值元素作为正确答案，问题为“实体的属性是？”的情形下，可以利用属性值集合约束信息、属性集合约束信息或属性值集合和属性集合约束信息三种。对于属性值集合约束信息，可以通过查找生成该问题的三元数据组中的属性值所对应的属性值的集合和生成该问题的三元数据组中的实体的属性对应的属性值的集合的差集，利用共现分析辅助生成。也就是说，从上述属性值的差集中选取与作为正确答案的属性值相似的选项作为错误选项，可选择多个这样的错误选项。例如，具有下述三元数据组：＜男明星 A、妻子、女明星 a＞＜男明星 B、妻子、女明星 b＞，＜男明星 C、妻子、女明星 c＞，＜男明星 D、妻子、女明星 d＞；对于问题：男明星 A 的妻子？，生成该问题的三元数据组＜男明星 A、妻子、女明星 a＞中的属性值对应的集合为 S1，即｛女明星 a｝，生成该问题的实体的属性对应的属性值的集合为 S2，即｛女明星 a，女明星 b，女明星 c，女明星 d｝。集合 S1 和 S2 的差集为｛女明星 b，女明星 c，女明星 d｝。利用这样的差集可生成相似错误选项：女明星 b，女明星 c，女明星 d。

[0053] 对于属性集合约束信息，可以利用夫妻属性、出生日期属性等作为约束信息。。例如，在属性为妻子或丈夫的情况下，可以限定错误选项为女性或男性。又例如，在属性为出生日期的情况下，可以限定出生日期大于 1900 等。

[0054] 对于热门程度（例如，可通过用户搜索量或点击量等进行量化）约束信息，可以用于对上面获取的相似的错误选项进行筛选，筛选掉热门程度较低的错误选项。在一些实施方式中，可以省略上面的相似选项选取步骤，仅利用热门程度生成热门程度较高的错误选项。

[0055] 对于公知常识约束信息，例如，在实体为“XX 妹妹”的情况下，她的性别属性，根据公知常识可知为女性，她的年龄不会是 40 岁以上。

[0056] 在本发明的一些实施方式，基于知识库的问答题目生成装置 200 还可以包括配图模块，该配图模块对正确答案和 / 或错误答案进行配图。通过上述的描述可知，正确答案或错误答案可以是实体、属性或属性值中的一种，对正确答案和 / 或错误的答案的配图可包括对实体、属性或属性值的配图。对于实体的配图，例如，对于人物实体，可以采用该人物的头像、剧照等图片。对于属性值（例如，星座、国籍、属相、职业、民族等）的配图，例如，对于国籍的配图，可以采用该国的国旗等图片。

[0057] 在本发明的实施方式中，基于知识库的问答题目生成装置 200 还可包括查看模块，该查看模块在生成的问题的显示页面呈现正确答案查看链接，供用户查看该问题的正确答案。例如，可以如图 2 所示，在问题的后面设置“偷看答案”选项，作为正确答案的查看链接，点击该连接，用户可查看该问题的正确答案，方便了用户实时了解自己的答题状况。

[0058] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本发明可借助软件结合硬件平台的方式来实现，当然也可以全部通过硬件来实施。基于这样的理解，本发明的技术方案对背景技术做出贡献的全部或者部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，智能手机或者网络设备等）执行本发明各个实施例或者实施例的某些部分所述的方法。

[0059] 本发明说明书中使用的术语和措辞仅仅为了举例说明，并不意味构成限定。本领域技术人员应当理解，在不脱离所公开的实施方式的基本原理的前提下，对上述实施方式中的各细节可进行各种变化。因此，本发明的范围只由权利要求确定，在权利要求中，除非另有说明，所有的术语应按最宽泛合理的意思进行理解。

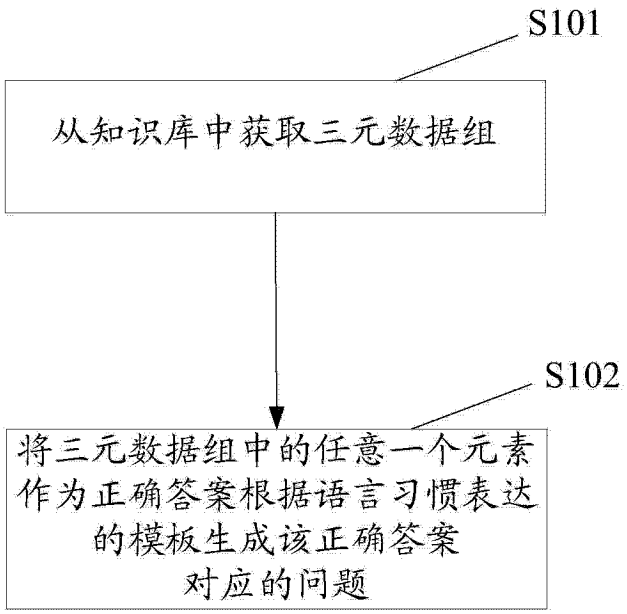

S101
从知识库中获取三元数据组
S102
将三元数据组中的任意一个元素
作为正确答案根据语言习惯表达
的模板生成该正确答案
对应的问题

图 1

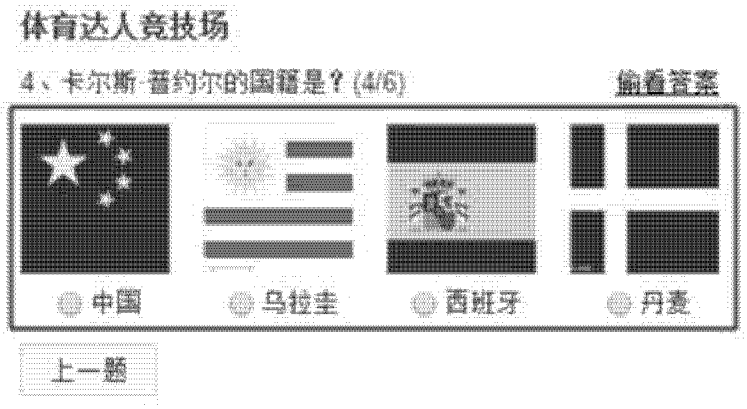

体育达人竞技场
4、卡尔斯·普约尔的国籍是？(4/6)
偷看答案
中国
乌拉圭
西班牙
丹麦
上一题

图 2

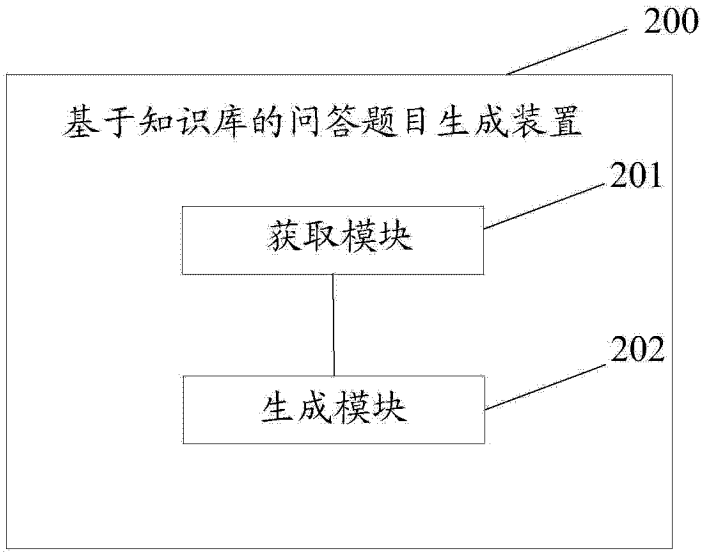


图 3