



(12) 发明专利申请

(10) 申请公布号 CN 103577558 A

(43) 申请公布日 2014. 02. 12

(21) 申请号 201310495881. 4

(22) 申请日 2013. 10. 21

(71) 申请人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街  
28号D座112室(德胜园区)

申请人 奇智软件(北京)有限公司

(72) 发明人 孙林 陈培军 秦吉胜

(74) 专利代理机构 北京市隆安律师事务所  
11323

代理人 权鲜枝

(51) Int. Cl.

G06F 17/30(2006. 01)

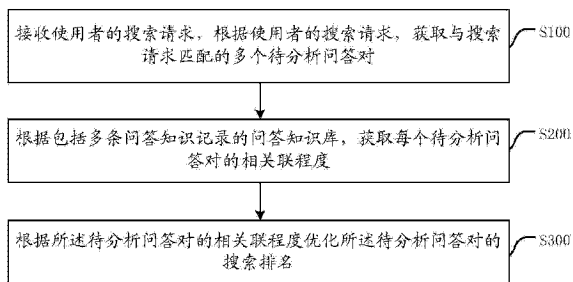
权利要求书2页 说明书16页 附图4页

(54) 发明名称

一种优化问答对的搜索排名的装置和方法

(57) 摘要

本发明公开了一种优化问答对的搜索排名的装置和方法,用于优化问答对搜索的搜索结果的排名,该方法包括如下步骤:接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问答对;根据包括多条问答知识记录的问答知识库,获取每个待分析问答对的相关联程度;根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。使用本发明的装置和方法,可以评价作为搜索结果的待分析问答对的相关联程度优化搜索结果排名,排名效果更好。



1. 一种优化问答对的搜索排名的装置,该装置包括:  
问答知识库,适于存储多条问答知识记录;  
搜索单元,适于接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问答对;  
相关联程度计算单元,适于根据问答知识库获取每个待分析问答对的相关联程度;  
搜索排名单元,适于根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。
2. 根据权利要求1所述的装置,其中,所述相关联程度计算单元包括:  
词语提取子单元,适于对待分析的问答对的问题内容和答案内容进行词语提取操作,得到至少一个待分析问题词语和至少一个待分析答案词语;  
计算子单元,适于根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算待分析的问答对的相关联程度。
3. 根据权利要求1或2所述的装置,其中,  
所述搜索排名单元,适于以所述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名。
4. 根据权利要求1至3任一项所述的装置,其中,该装置还包括问答知识库构建单元,所述问答知识库构建单元,适于预先从含有问答对的网页提取出多个问答对,根据提取的问答对构建包括多条问答知识记录的问答知识库;  
所述问答知识库构建单元,进一步适于在从含有问答对的网页提取出多个问答对时,抓取与所述问答对对应的类别;  
所述问答知识库构建单元,进一步适于在根据提取的问答对构建问答知识库时,根据问答对和与所述问答对对应的类别构建问答知识记录;每个问答知识记录对应于一个类别,分别包括一个问题词语、一个答案词语,以及所述问题词语和所述答案词语之间的语义相关度。
5. 根据权利要求1至4任一项所述的装置,其中,  
所述计算子单元,适于选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录;根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对针对各个类别的相关联程度;选取上述该待分析的问答对针对各个类别的相关联程度的最大值,以该最大值作为待分析的问答对的相关联程度。
6. 一种优化问答对的搜索排名的方法,该方法包括如下步骤:  
接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问答对;  
根据包括多条问答知识记录的问答知识库,获取每个待分析问答对的相关联程度;  
根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。
7. 根据权利要求6所述的方法,其中,所述根据包括多条问答知识记录的问答知识库获取每个待分析问答对的相关联程度,包括对每个待分析问答对执行以下操作:  
对该待分析的问答对的问题内容和答案内容进行词语提取操作,得到至少一个待分析问题词语和至少一个待分析答案词语;

根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算该待分析的问答对的相关联程度。

8. 根据权利要求6或7所述的方法,其中,所述根据所述待分析问答对的相关联程度调整所述待分析问答对的搜索排名,具体包括:

以所述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名。

9. 根据权利要求6至8任一项所述的方法,其中,该方法进一步包括:

预先从含有问答对的网页提取出多个问答对,根据提取的问答对构建包括多条问答知识记录的问答知识库;

在从含有问答对的网页提取出多个问答对时,抓取与所述问答对对应的类别;

在根据提取的问答对构建问答知识库时,根据问答对和与所述问答对对应的类别构建问答知识记录;

每个问答知识记录对应于一个类别,分别包括一个问题词语、一个答案词语,以及所述问题词语和所述答案词语之间的语义相关度。

10. 根据权利要求6至9任一项所述的方法,其中,

所述根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算待分析的问答对的相关联程度,具体包括:

选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录;

根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对针对各个类别的相关联程度;

选取上述该待分析的问答对针对各个类别的相关联程度的最大值,以该最大值作为待分析的问答对的相关联程度。

## 一种优化问答对的搜索排名的装置和方法

### 技术领域

[0001] 本发明涉及网络数据通信领域,具体涉及一种优化问答对的搜索排名的装置和方法。

### 背景技术

[0002] 问答社区是一种用户产生内容的网络应用,基本形式是由用户根据自己的需求提出问题,并由其他的用户来给出回答。这种形式为用户在网络上获取信息提供了新的渠道。然而由于任何用户都可以随意地创建内容,导致了问答社区中的信息质量差异非常大,以至于问答社区中出现了大量的低质量问答对。这不但降低了问答社区的质量,更给用户查找信息带来了诸多不便,例如,使用现有的搜索技术进行问答搜索时,获取的搜索结果中存在部分低质量的问答对而现有技术对搜索结果进行排序的方法,更多地依赖于问答对所属的网站和问答对的非文本特征来对问答对进行排序,会影响精确性和通用性。

### 发明内容

[0003] 鉴于上述问题,提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种优化问答对的搜索排名的装置和相应的优化问答对的搜索排名的方法。

[0004] 依据本发明的一个方面,提供了一种优化问答对的搜索排名的装置,该装置包括:

[0005] 问答知识库,适于存储多条问答知识记录;

[0006] 搜索单元,适于接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问答对;相关联程度计算单元,适于根据问答知识库获取每个待分析问答对的相关联程度;

[0007] 搜索排名单元,适于根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。

[0008] 可选地,所述相关联程度计算单元包括:词语提取子单元,适于对待分析的问答对的问题内容和答案内容进行词语提取操作,得到至少一个待分析问题词语和至少一个待分析答案词语;计算子单元,适于根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算待分析的问答对的相关联程度。

[0009] 可选地,所述搜索排名单元,适于以所述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名;或,根据搜索排列技术初步排列所述待分析问答对所属的网站,根据将该初步排列的次序号与所述待分析问答对的相关联程度计算所述待分析问答对的搜索排名。

[0010] 可选地,该装置还包括问答知识库构建单元,所述问答知识库构建单元,适于预先从含有问答对的网页提取出多个问答对,根据提取的问答对构建包括多条问答知识记录的问答知识库;所述问答知识库构建单元,进一步适于在从含有问答对的网页提取出多个问答对时,抓取与所述问答对对应的类别;所述问答知识库构建单元,进一步适于在根据提取

的问答对构建问答知识库时,根据问答对和与所述问答对对应的类别构建问答知识记录;每个问答知识记录对应于一个类别,分别包括一个问题词语、一个答案词语,以及所述问题词语和所述答案词语之间的语义相关度。可选地,所述计算子单元,适于选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录;根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对针对各个类别的相关联程度;选取上述该待分析的问答对针对各个类别的相关联程度的最大值,以该最大值作为待分析的问答对的相关联程度。

[0011] 可选地,所述计算子单元,适于将选取的问答知识记录中对应于相同类别的问答知识记录的语义相关度加权相加,得到该待分析的问答对分别针对各个类别的相关联程度。

[0012] 可选地,所述词语提取子单元,适于对待分析的问答对的问题内容和答案内容进行分词、去除停用词、词合并,和提取实体词的操作。

[0013] 可选地,所述问答知识库构建单元,适于对每个问答对执行以下操作:对该问答对的问题内容和答案内容进行词语提取操作,得到问题词语集合和答案词语集合;令问题词语集合中的每个问题词语与答案词语集合中的每个答案词语分别在与该问答对对应的每个类别上形成一条信息记录;所述问答知识库构建单元,适于对每一条信息记录,执行以下操作:计算该答案词语属于该类别的概率,计算在该类别上该答案词语对该问题词语的解释的专一程度,计算在该类别上该问题词语用该答案词语进行解释的强度;将上述概率、专一程度和强度相乘,所得到的乘积是该答案词语和该问题词语的语义相关度;令该问题词语、该答案词语和其语义相关度形成一条对应于该类别的问答知识记录。

[0014] 可选地,所述问答知识库构建单元,适于按照如下的方法计算该答案词语属于该类别的概率:

$$[0015] \quad P(Ck|AWj) = P(AWj|Ck) * \frac{P(Ck)}{P(AWj)}$$

[0016] 所述问答知识库构建单元,适于按照如下的方法计算在该类别上各个答案词语对该问题词语的解释的专一程度:

$$[0017] \quad \text{specific}(QWi, AWj|C = Ck) = P(QWi|AWj, C = Ck) = \frac{\#(QWi, AWj)}{\#(AWj)}|C = Ck;$$

[0018] 所述问答知识库构建单元,适于按照如下的方法计算在该类别上该问题词语用各个答案词语进行解释的强度:

$$[0019] \quad \text{interpret}(QWi, AWj|C = Ck) = P(AWj|QWi, C = Ck) = \frac{\#(QWi, AWj)}{\sum_{j=1}^x \#(QWi, AWj)}|C = Ck;$$

[0020] 所述问答知识库构建单元,适于按照如下的方法将上述概率、专一程度和强度相乘:

$$[0021] \quad \text{weight}(QWi, AWj|C = Ck) = P(Ck|AWj) * \text{specific}(QWi, AWj|C = Ck) * \text{interpret}(QWi, AWj|C = Ck);$$

[0022] 其中,  $P(Ck)$  表示类别  $Ck$  出现的概率;  $P(AWj)$  表示答案为  $AWj$  的概率;  $P(AWj | Ck)$  表示  $Ck$  类别属于  $AWj$  的概率;

[0023] # (QW<sub>i</sub>, AW<sub>j</sub>) 表示问题词语为 QW<sub>i</sub> 且答案词语为 AW<sub>j</sub> 的次数；

[0024] # (AW<sub>j</sub>) 表示答案词语为 AW<sub>j</sub> 的次数。

[0025] 根据本发明的另一方面,提供了一种优化问答对的搜索排名的方法,该方法包括如下步骤:

[0026] 接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问答对;

[0027] 根据包括多条问答知识记录的问答知识库获取每个待分析问答对的相关联程度;

[0028] 根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。

[0029] 可选地,所述根据包括多条问答知识记录的问答知识库优化每个待分析问答对的相关联程度,包括对每个待分析问答对执行以下操作:对该待分析的问答对的问题内容和答案内容进行词语提取操作,得到至少一个待分析问题词语和至少一个待分析答案词语;根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算该待分析的问答对的相关联程度。

[0030] 可选地,所述根据所述待分析问答对的相关联程度调整所述待分析问答对的搜索排名,具体包括:以所述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名;或,根据搜索排列技术初步排列所述待分析问答对所属的网站,根据该初步排列的次序号与所述待分析问答对的相关联程度计算所述待分析问答对的搜索排名。

[0031] 可选地,该方法进一步包括:预先从含有问答对的网页提取出多个问答对,根据提取的问答对构建包括多条问答知识记录的问答知识库;在从含有问答对的网页提取出多个问答对时,抓取与所述问答对对应的类别;在根据提取的问答对构建问答知识库时,根据问答对和与所述问答对对应的类别构建问答知识记录;每个问答知识记录对应于一个类别,分别包括一个问题词语、一个答案词语,以及所述问题词语和所述答案词语之间的语义相关度。

[0032] 可选地,所述根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算待分析的问答对的相关联程度,具体包括:选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录;根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对针对各个类别的相关联程度;选取上述该待分析的问答对针对各个类别的相关联程度的最大值,以该最大值作为待分析的问答对的相关联程度。

[0033] 可选地,根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对分别针对各个类别的相关联程度,具体包括:将选取的问答知识记录中对应于相同类别的问答知识记录的语义相关度加权相加,得到该待分析的问答对分别针对各个类别的相关联程度。

[0034] 可选地,所述对所述待分析的问答对的问题内容和答案内容进行词语提取操作,具体包括:对待分析的问答对的问题内容和答案内容进行分词、去除停用词、词合并,和提取实体词的操作。

[0035] 可选地,所述根据问答对和与所述问答对对应的类别构建问答知识库,具体包括:对每个问答对,对该问答对的问题内容和答案内容进行词语提取操作,得到问题词语集合

和答案词语集合；令问题词语集合中的每个问题词语与答案词语集合中的每个答案词语分别在与其对应的每个类别上形成一条信息记录；对每一条信息记录，执行以下操作：计算该答案词语属于该类别的概率，计算在该类别上该答案词语对该问题词语的解释的专一程度，计算在该类别上该问题词语用该答案词语进行解释的强度；将上述概率、专一程度和强度相乘，所得到的乘积是该答案词语和该问题词语的语义相关度；令该问题词语、该答案词语和其语义相关度形成一条对应于该类别的问答知识记录。

[0036] 可选地，所述计算该答案词语属于该类别的概率，具体包括：

$$[0037] \quad P(Ck|AWj) = P(AWj|Ck) * \frac{P(Ck)}{P(AWj)};$$

[0038] 所述计算在该类别上各个答案词语对该问题词语的解释的专一程度，具体包括：

$$[0039] \quad specific(QWi, AWj|C = Ck) = P(QWi|AWj, C = Ck) = \frac{\#(QWi, AWj)}{\#(AWj)}|C = Ck;$$

[0040] 所述计算在该类别上该问题词语用各个答案词语进行解释的强度，具体包括：

$$[0041] \quad interpret(QWi, AWj|C = Ck) = P(AWj|QWi, C = Ck) = \frac{\#(QWi, AWj)}{\sum_{j=1}^x \#(QWi, AWj)}|C = Ck;$$

[0042] 将上述概率、专一程度和强度相乘，具体包括：

$$[0043] \quad weight(QWi, AWj|C = Ck) = P(Ck|AWj) * specific(QWi, AWj|C = Ck) * interpret(QWi, AWj|C = Ck);$$

[0044] 其中， $P(Ck)$ 表示类别  $Ck$  出现的概率； $P(AWj)$ 表示答案为  $AWj$  的概率； $P(AWj | Ck)$ 表示  $Ck$  类别属于  $AWj$  的概率；

[0045]  $\#(QWi, AWj)$ 表示问题词语为  $QWi$  且答案词语为  $AWj$  的次数；

[0046]  $\#(AWj)$ 表示答案词语为  $AWj$  的次数。

[0047] 根据本发明的技术方案，从含有问答对的网页提取出多个问答对并根据提取的问答对构建包括多条问答知识记录的问答知识库，根据使用者的搜索请求获取的与搜索请求匹配的多个待分析问答对，根据问答知识库获取每个待分析问答对的相关联程度并根据待分析问答对的相关联程度优化待分析问答对的搜索排名，可以从语义方面评价待分析问答对的质量，解决了现有技术依赖于问答对所属的网页和问答对的非文本特征来对问答对进行排序而导致的排序效果不佳的问题，而且容易实现、通用性强。

## 附图说明

[0048] 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

[0049] 图 1 示出了根据本发明一个实施例的优化问答对的搜索排名的方法的流程图；

[0050] 图 2 示出了构建问答知识库的详细的流程图；

[0051] 图 3 示出了使用如图 2 所示的步骤而得到的问答知识库的一个解释模型示意图；

[0052] 图 4 示出了图 1 中步骤 S200 的详细的流程图；

[0053] 图 5 示出了图 4 中步骤 S220 的详细的流程图；以及

- [0054] 图 6 示出了根据本发明一个实施例的优化问答对的搜索排名的装置的框图；
- [0055] 图 7 示出了图 6 中相关联程度计算单元 300 的详细的框图；
- [0056] 图 8 示出了根据本发明另一个实施例的优化问答对的搜索排名的装置的框图。

### 具体实施方式

[0057] 现有的获取问答对的搜索排名的方法，是使用文本特征和非文本特征来描述问答对的问题和答案从而对问答对进行排名，或根据问答对所属的网站的排名对问答对进行排名。文本特征主要包括文本视觉特征(例如标点符号密度，平均词长，文本熵等)和文本内容特征(例如文本内容词比例，疑问词密度，相关词覆盖等)，并提取中文自动差错广泛采用的特征(例如单字密度特征等)；非文本特征包含用户的权威度指标，答案问题状态，答案回答时间，用户关系交互特征等。在对问题和答案分别提取出特征后，在训练集上分别学习出一个问题质量预测模型和答案质量预测模型，并使用两个模型的输出结果来评价问答对质量。然而，使用现有的获取问答对的相关联程度的方法对于答案质量进行评价时，仅仅使用了相关词覆盖特征来描述问题和答案间的语义匹配度，这不但仅仅是停留在词法层面上的，而且没有考虑问题和答案间的语义匹配度。然而问题和答案间的语义匹配度恰恰是问答对质量的核心，比如问题为“中国的首都是哪里？”，答案 1 为“北京”，答案 2 为“中国的首都是上海”。那么问题经过分词及丢弃停用词处理后，为“中国首都哪里”，答案 1 分词结果为“北京”，答案 2 分词结果为“中国首都上海”。现有技术中，语义匹配度可以定义为：问题和答案中共同出现的词语个数除以问题和答案中所有词语的个数。则问题和答案 1 的语义匹配度为： $0/4=0$ 。问题和答案 2 的语义匹配度为： $2/4=0.5$ 。使用现有技术，就会认为答案 2 和问题较为匹配，从而答案 2 对应的问答对在搜索结果(例如，当使用者的搜索条件为“首都”，或者“中国首都”等)中的排名往往在前。而我们知道这显然是不当的。

[0058] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

[0059] 图 1 示出了根据本发明一个实施例的优化问答对的搜索排名的方法的流程图。该方法包括如下步骤 S100、步骤 S200 和步骤 S300：

[0060] S100、接收使用者的搜索请求，根据使用者的搜索请求，获取与搜索请求匹配的多个待分析问答对。

[0061] 在本发明的一个实施例中，可以是使用网络搜索技术，例如使用问答对搜索引擎，根据使用者的搜索请求获取待分析问答对。

[0062] S200、根据包括多条问答知识记录的问答知识库，获取每个待分析问答对的相关联程度。

[0063] 本实施例的步骤 S200，可以通过利用问答知识库从语义方面对待分析问答对的问题内容和答案内容进行分析以获得待分析问答对的相关联程度，评价效果更好而且容易实现。

[0064] 进一步地，所述包括多条问答知识记录的问答知识库，是通过预先从含有问答对的网页提取出多个问答对，根据提取的问答对构建而得到的。在本发明的一个实施例中，



在从含有问答对的网页提取出多个问答对时,抓取与所述问答对对应的类别。则在根据提取的问答对构建问答知识库时,根据问答对和与所述问答对对应的类别构建问答知识记录。得到的问答知识库之中的每个问答知识记录对应于一个类别,分别包括一个问题词语(QW)、一个答案词语(AW),以及所述问题词语和所述答案词语之间的语义相关度。通过利用由网页提取的海量的、高质量的问答对构建包括多条问答知识记录的问答知识库,可以基于对海量信息的学习而获取多条问答知识记录的问题词语和答案词语之间的语义相关度;通过利用从网页提取得到的信息构建问答知识库,适用的范围更广,方法的通用性更强。

[0065] S300、根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。

[0066] 由于待分析问答对的相关联程度反映了质量,所以可以利用相关联程度优化所述待分析问答对的搜索排名,排名效果更好。

[0067] 具体的方法,可以是以所述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名,即相关联程度高的问答对的搜索排名靠前;也可以是先根据搜索排列技术初步排列所述待分析问答对所属的网站,根据该初步排列的次序号与所述待分析问答对的相关联程度计算所述待分析问答对的搜索排名,例如,可以将所述待分析问答对所属的网站的初步排列的次序号与所述待分析问答对的相关联程度相乘,以相乘运算的结果的次序作为所述待分析问答对的搜索排名;通过将待分析问答对的质量和其所属网站的排名结合,以对待分析问答对进行排序,使用者使用问答对搜索时,能够获得更好的结果排序的质量。

[0068] 图 2 示出了构建问答知识库的详细的流程图。具体包括以下步骤 S410、步骤 S420 和步骤 S430:

[0069] S410、预先从含有问答对的网页提取出多个问答对,抓取与所述问答对对应的类别。

[0070] 本实施例中,可以通过使用网络爬虫,从互联网上含有高质量问答对的网页抓取数据并提取出问答对,以保证所提取的问答对的质量;所述含有高质量问答对的网页包括 cQA 社区、各大专业论坛等,则可以使用楼层识别技术,根据楼主提问题,1 楼 2 楼等为答案的方式,来提取问答对。由于所述含有高质量问答对的网页中包括对应于每个问答对的类别信息,所以可以在抓取问答对的同时一并抓取与所述问答对对应的类别。

[0071] S420、对每个问答对,对该问答对的问题内容和答案内容进行词语提取操作,得到问题词语集合和答案词语集合;令问题词语集合中的每个问题词语与答案词语集合中的每个答案词语分别在与该问答对对应的每个类别上形成一条信息记录。

[0072] 在本发明的一个实施例中,对步骤 S410 中提取得到的所述问答对中的每一个问答对的问题内容和答案内容进行词语提取操作,具体包括,对问答对的问题内容和答案内容进行分词、去除停用词、词合并,和提取实体词的操作。

[0073] 则由每一个问答对的问题内容得到至少一个问题词语,由每一个问答对的答案内容得到至少一个答案词语,则可以得到针对该问答对的类别集合  $\langle C_1, \dots, C_k, \dots, C_p \rangle$ 、问题词语集合  $\langle QW_1, \dots, QW_i, \dots, QW_m \rangle$  和答案词语集合  $\langle AW_1, \dots, AW_j, \dots, AW_n \rangle$ 。

[0074] 通过令问题词语集合中的每个问题词语( $QW_i$ )与答案词语集合中的每个答案词语( $AW_j$ )分别在与该问答对对应的每个类别( $C_k$ )上形成一条信息记录,例如  $\langle QW_i, AW_j, C_k \rangle$ ,则可以形成  $m*n*p$  条信息记录。

[0075] S430、对每一条信息记录,执行以下操作:计算该答案词语属于该类别的概率,计算在该类别上该答案词语对该问题词语的解释的专一程度,计算在该类别上该问题词语用该答案词语进行解释的强度;将上述概率、专一程度和强度相乘,所得到的乘积是该答案词语和该问题词语的语义相关度;令该问题词语、该答案词语和其语义相关度形成一条对应于该类别的问答知识记录  $\langle QW_i, AW_j, \text{weight}(QW_i, AW_j) \rangle$  或  $\langle QW_i, AW_j, C_k, \text{weight}(QW_i, AW_j) \rangle$ 。本实施例中的步骤 S430,可以是在对从网页抓取的海量的问答对进行了如步骤 S420 所述的词语提取操作而得到海量的信息记录之后基于所述海量的信息记录进行的,则基于海量的信息记录而获取的语义相关度更准确。

[0076] 较佳地,所述计算该答案词语属于该类别的概率,具体包括:

$$[0077] \quad P(C_k | AW_j) = P(AW_j | C_k) * \frac{P(C_k)}{P(AW_j)}$$

[0078] 所述计算在该类别上各个答案词语对该问题词语的解释的专一程度,具体包括:

$$[0079] \quad \text{specific}(QW_i, AW_j | C = C_k) = P(QW_i | AW_j, C = C_k) = \frac{\#(QW_i, AW_j)}{\#(AW_j)} | C = C_k;$$

[0080] 所述计算在该类别上该问题词语用各个答案词语进行解释的强度,具体包括:

$$[0081] \quad \text{interpret}(QW_i, AW_j | C = C_k) = P(AW_j | QW_i, C = C_k) = \frac{\#(QW_i, AW_j)}{\sum_{j=1}^x \#(QW_i, AW_j)} | C = C_k;$$

[0082] 将上述概率、专一程度和强度相乘,具体包括:

$$[0083] \quad \text{weight}(QW_i, AW_j | C = C_k) = P(C_k | AW_j) * \text{specific}(QW_i, AW_j | C = C_k) * \text{interpret}(QW_i, AW_j | C = C_k);$$

[0084] 其中,  $P(C_k)$  表示类别  $C_k$  出现的概率;  $P(AW_j)$  表示答案为  $AW_j$  的概率;  $P(AW_j | C_k)$  表示  $C_k$  类别属于  $AW_j$  的概率;

[0085]  $\#(QW_i, AW_j)$  表示问题词语为  $QW_i$  且答案词语为  $AW_j$  的次数;

[0086]  $\#(AW_j)$  表示答案词语为  $AW_j$  的次数。

[0087] 由步骤 S410、步骤 S420 和步骤 S430,可以得到问答知识记录而构建问答知识库。图 3 示出了使用如图 2 所示的步骤而得到的问答知识库的一个解释模型示意图。可知,对于每一问题词语  $QW_i$ ,可以针对类别集合  $\langle C_1, \dots, C_k, \dots, C_p \rangle$  中的每一类别,获得  $n$  条问答知识记录。当然,本领域技术人员可以了解的是,若计算得到的语义相关度为 0,则可以删除相应的问答知识记录;再者,如果问答知识库中问答知识记录的数量过大而使得存储问答知识记录和计算待分析问答对的相关联程度的开销过大,可以预设一个阈值,将语义相关度小于阈值的问答知识记录删除以减小开销。

[0088] 图 4 示出了图 1 中步骤 S200 的详细流程图。步骤 S200 具体包括以下步骤 S210 和步骤 S220。

[0089] S210、对待分析的问答对的问题内容和答案内容进行词语提取操作,得到至少一个待分析问题词语和至少一个待分析答案词语。

[0090] 在本发明的一个实施例中,对待分析的问答对的问题内容和答案内容进行词语提取操作具体包括:对待分析的问答对的问题内容和答案内容进行分词、去除停用词、词合并(word join),和提取实体词(例如名词、动词等)的操作。则由待分析的问答对的问题内容

得到至少一个待分析问题词语,由待分析的问答对的答案内容得到至少一个待分析答案词语。

[0091] S220、根据待分析问题词语和待分析答案词语,从问答知识库选择至少一条问答知识记录,根据所选择的问答知识记录计算待分析的问答对的相关联程度。

[0092] 图 5 示出了图 4 中步骤 S220 的详细流程图。在通过步骤 S210 得到至少一个待分析问题词语和至少一个待分析答案词语后,步骤 S220 具体包括以下步骤 S221、步骤 S222 和步骤 S223:

[0093] S221、选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录。本实施例中,问题词语与待分析问题词语匹配是指待分析问题词语与问题词语相同或待分析问题词语是问题词语的子串;答案词语与待分析答案词语匹配是指待分析答案词语与答案词语相同或待分析答案词语是答案词语的子串,本实施例通过步骤 S210,使用字段匹配或字段搜索的方法,从问答知识库中选出部分与待分析的问答对相关的问答知识记录。

[0094] S222、根据所述选取的问答知识记录中对应于相同类别的问答知识记录,得到该待分析的问答对分别针对各个类别的相关联程度,具体包括:将选取的问答知识记录中对应于相同类别的问答知识记录的语义相关度加权相加,得到该待分析的问答对分别针对各个类别的相关联程度。

[0095] 本实施例,将通过步骤 S221 选出的问答知识记录根据其所对应的类别进行分组,对应于相同类别的问答知识记录为一组;将每一组的问答知识记录的语义相关度加权(例如,权值为 1 或 100)相加,得到该待分析的问答对针对该类别的相关联程度;由此得到至少一个(本实施例中的相关关联程度的数目即待分析问题对对应的类别的数目)相关关联程度。

[0096] S223、选取上述该待分析的问答对针对各个类别的相关关联程度的最大值,以该最大值作为待分析的问答对的相关关联程度。

[0097] 图 6 示出了根据本发明一个实施例的优化问答对的搜索排名的装置的框图。该装置包括问答知识库 100、搜索单元 200、相关关联程度计算单元 300 和搜索排名单元 400。

[0098] 问答知识库 100,适于存储多条问答知识记录。本实施例的问答知识库 100 能够通过抓取网页中的海量问答对构建得到。

[0099] 搜索单元 200,适于接收使用者的搜索请求,根据使用者的搜索请求,获取与搜索请求匹配的多个待分析问题对。

[0100] 在本发明的一个实施例中,搜索单元 200 可以是问答对搜索引擎,根据使用者的搜索请求获取待分析问题对;例如搜索单元 200 是用于问答对搜索的网络搜索引擎,接收使用者通过浏览器输入的搜索请求并获取待分析问题对。

[0101] 相关关联程度计算单元 300,适于根据问答知识库获取每个待分析问题对的相关关联程度。

[0102] 本发明的相关关联程度计算单元 300 可以通过利用问答知识库从语义方面对待分析问题对的问题内容和答案内容进行分析以获得待分析问题对的相关关联程度,评价效果更好而且容易实现。问答知识库 100 利用由网页提取的海量的、高质量的问答对构建并且包括多条问答知识记录,可以基于对海量信息的学习而获取多条问答知识记录的问题词语和答案词语之间的语义相关度。

[0103] 搜索排名单元 400, 适于根据所述待分析问答对的相关联程度优化所述待分析问答对的搜索排名。

[0104] 由于待分析问答对的相关联程度反映了质量, 所以可以利用相关联程度优化所述待分析问答对的搜索排名, 排名效果更好。具体的方法, 可以是以下述待分析问答对的相关联程度的次序作为所述待分析问答对的搜索排名, 即相关联程度高的问答对的搜索排名靠前; 也可以是先根据搜索排列技术初步排列所述待分析问答对所属的网站, 根据该初步排列的次序号与所述待分析问答对的相关联程度计算所述待分析问答对的搜索排名, 例如, 可以将所述待分析问答对所属的网站的初步排列的次序号与所述待分析问答对的相关联程度相乘, 以相乘运算的结果的次序作为所述待分析问答对的搜索排名。

[0105] 图 7 示出了图 6 中相关联程度计算单元 300 的详细的框图。相关联程度计算单元 300 包括词语提取子单元 310 和计算子单元 320。

[0106] 词语提取子单元 310, 适于对待分析的问答对的问题内容和答案内容进行词语提取操作, 得到至少一个待分析问题词语和至少一个待分析答案词语。

[0107] 在本发明的一个实施例中, 词语提取子单元 310, 适于对待分析的问答对的问题内容和答案内容进行分词、去除停用词、词合并(word join), 和提取实体词(例如名词、动词等)的操作, 以得到至少一个待分析问题词语和至少一个待分析答案词语。

[0108] 计算子单元 320, 适于根据待分析问题词语和待分析答案词语, 从问答知识库选择至少一条问答知识记录, 根据所选择的问答知识记录计算待分析的问答对的相关联程度。

[0109] 在本发明的一个实施例中, 计算子单元 320, 适于选取其包括的问题词语与待分析问题词语匹配且包括的答案词语与待分析答案词语匹配的问答知识记录。本实施例中, 问题词语与待分析问题词语匹配是指待分析问题词语与问题词语相同或待分析问题词语是问题词语的子串; 答案词语与待分析答案词语匹配是指待分析答案词语与答案词语相同或待分析答案词语是答案词语的子串; 根据所述选取的问答知识记录中对应于相同类别的问答知识记录, 得到该待分析的问答对针对各个类别的相关联程度, 更具体地, 是将选取的问答知识记录中对应于相同类别的问答知识记录的语义相关度加权(例如, 权值为 1 或 100)相加而得到该待分析的问答对分别针对各个类别的相关联程度, 由此得到至少一个(本实施例中的相关联程度的数目即待分析问答对对应的类别的数目)相关联程度; 选取上述该待分析的问答对针对各个类别的相关联程度的最大值, 以该最大值作为待分析的问答对的相关联程度。

[0110] 图 8 示出了根据本发明另一个实施例的确定网络资源点的抓取频率的装置的框图。在本实施例中, 该装置还包括问答知识库构建单元 500, 问答知识库构建单元 500 适于预先从含有问答对的网页提取出多个问答对, 根据提取的问答对构建包括多条问答知识记录的问答知识库。在图 6 所示的装置中, 问答知识库是已有的, 由于实际网络的信息量不断增加, 信息内容的变化速度快, 问答知识库的内容往往需要更新, 本实施例通过增设问答知识库构建单元 500 构建(或者说更新)问答知识库, 可以保证问答知识库的内容的即时性和可靠性。

[0111] 较佳地, 在从含有问答对的网页提取出多个问答对时, 问答知识库构建单元 500 抓取与所述问答对对应的类别。本实施例中, 可以通过使用网络爬虫, 从互联网上含有高质量问答对的网页抓取数据并提取出问答对, 以保证所提取的问答对的质量; 所述含有高质

量问答对的网页包括 cQA 社区、各大专业论坛等。由于所述含有高质量问答对的网页中包括对应于每个问答对的类别信息,所以问答知识库构建单元 500 可以在抓取问答对的同时一并抓取与所述问答对对应的类别。

[0112] 在本实施例中,问答知识库构建单元 500,适于对每个问答对执行以下操作:对该问答对的问题内容和答案内容进行词语提取操作,得到问题词语集合和答案词语集合,具体地,问答知识库构建单元 500 对提取得到的所述问答对中的每一个问答对的问题内容和答案内容进行分词、去除停用词、词合并,和提取实体词的操作而得到问题词语和答案词语;令问题词语集合中的每个问题词语与答案词语集合中的每个答案词语分别在与该问答对对应的每个类别上形成一条信息记录。问答知识库构建单元 500,适于对每一条信息记录,执行以下操作:计算该答案词语属于该类别的概率,计算在该类别上该答案词语对该问题词语的解释的专一程度,计算在该类别上该问题词语用该答案词语进行解释的强度;将上述概率、专一程度和强度相乘,所得到的乘积是该答案词语和该问题词语的语义相关度;令该问题词语、该答案词语和其语义相关度形成一条对应于该类别的问答知识记录。

[0113] 更具体地,问答知识库构建单元 500,适于按照如下的方法计算该答案词语属于该类别的概率:

$$[0114] \quad P(Ck|AWj) = P(AWj|Ck) * \frac{P(Ck)}{P(AWj)};$$

[0115] 更具体地,问答知识库构建单元 500,适于按照如下的方法计算在该类别上各个答案词语对该问题词语的解释的专一程度:

$$[0116] \quad \text{specific}(QWi, AWj|C = Ck) = P(QWi|AWj, C = Ck) = \frac{\#(QWi, AWj)}{\#(AWj)}|C = Ck;$$

[0117] 更具体地,问答知识库构建单元 500,适于按照如下的方法计算在该类别上该问题词语用各个答案词语进行解释的强度:

$$[0118] \quad \text{interpret}(QWi, AWj|C = Ck) = P(AWj|QWi, C = Ck) = \frac{\#(QWi, AWj)}{\sum_{j=1}^x \#(QWi, AWj)}|C = Ck;$$

[0119] 更具体地,问答知识库构建单元 500,适于按照如下的方法将上述概率、专一程度和强度相乘:

$$[0120] \quad \text{weight}(QWi, AWj|C = Ck) = P(Ck|AWj) * \text{specific}(QWi, AWj|C = Ck) * \text{interpret}(QWi, AWj|C = Ck);$$

[0121] 其中,  $P(Ck)$  表示类别  $Ck$  出现的概率;  $P(AWj)$  表示答案为  $AWj$  的概率;  $P(AWj | Ck)$  表示  $Ck$  类别属于  $AWj$  的概率;

[0122]  $\#(QWi, AWj)$  表示问题词语为  $QWi$  且答案词语为  $AWj$  的次数;

[0123]  $\#(AWj)$  表示答案词语为  $AWj$  的次数。

[0124] 以下通过一个例子说明使用本发明的实施例所能达到的效果,比如有如下问答对,类别为“医疗健康”:

[0125]

问题	孩子咳嗽流鼻涕怎么办
答案	这是小儿上感的症状,可先给予药物治疗: 1、抗病毒: 小儿感冒颗粒,按说明剂量,日服3次; 2、止咳: 小儿咳喘灵(中药冲剂),按说明剂量,日服3次; 必要时加用抗生素: 阿莫西林颗粒(再林)口服。或用安克口服。也可用太儿欣(罗红霉素)口服。如药物口服疗效不佳时,应及时检查,或抗生素静滴治疗。

[0126] 通过分词技术处理,得到待分析问题词语和待分析答案词语如下:

[0127]

待分析问题词语	孩子, 咳嗽, 鼻涕
待分析答案词语	症状, 药物, 治疗, 抗病毒, 小儿感冒颗粒, 说明, 剂量, 止咳, 中药, 冲剂, 抗生素, 阿莫西林, 阿莫西林颗粒, 颗粒, 口服, 罗红霉素, 疗效

[0128] 从分词结果可以看出,问题和答案中没有相关词覆盖,因此如果使用现有技术则容易认为该问答对相关程度低,质量不高,因此搜索排名靠后。但是实际上使用人工判断明显可知该问答对是一个高质量的问答对。

[0129] 若使用本发明的方法和装置处理,首先,可以调取已有的问答知识库,或者通过抓取 cQA 社区、各大专业论坛的问答对,构建得到问答知识库;

[0130] 第二步,在接收使用者的搜索请求,根据使用者的搜索请求(例如,孩子鼻涕),获取与搜索请求匹配的多个待分析问答对,假定搜索结果中包括上述待分析的问答对;

[0131] 第三步,对上述待分析的问答对,经过词语提取操作得到待分析问题词语集合<孩子,咳嗽,鼻涕>、待分析答案词语集合<症状,药物,治疗,抗病毒,小儿感冒颗粒,说明,剂量,止咳,中药,冲剂,抗生素,阿莫西林,阿莫西林颗粒,颗粒,口服,罗红霉素,疗效>,并且得到待分析的问答对的类别为“医疗健康”;根据各个待分析问题词语以及该类别,从问答知识库中选择得到问题词语与待分析问题词语匹配的若干问答知识记录,从而得到如下答案词语及语义相关度(为了方便阅读,下表中的语义相关度的数值是进行了适当的归一化处理后的数值):

[0132]

问题词语	类别	答案词语及语义相关度
孩子	医疗健康	快克:32, 治疗:11, 感冒药:7, 检查:7, 月经:5, 病情:5, 妊娠:4, 患儿:4, 面诊患者:3, 排卵:3, 发烧:3, 剖腹产:3, 药流:2, 精子:2, 抽动症:2, 输卵管:2, 家长:2, 保胎:2, 发生:1, 休息:1, 了解:1, 恢复:1, 锌硒宝:1, 癫痫:1, 对症:1 病因:1, 筛查:1, 生命阳光:1, 上海康尧:1, 儿科:1, 复查:1, 哺乳:1, 调理:1 抵抗力:1, 食物:1, 人工流产:1, 脑瘫:1, 呕吐:1, 叶酸:1, 康尧:1, 苯丙醇:1, 神经:1
咳嗽	医疗健康	支气管炎:40, 婴儿久:19, 阴功效:18, 咽炎:18, 剑花猪肺汤:18, 糖婴儿:18, 慢性气管炎支气管哮喘:18, 丝瓜粥:17, 痰热:15, 猪肺:15, 南杏:15, 款冬花:13, 果猪肺汤:12, 杏仁热水提取物粗制剂:12, 润肺:12, 噬菌体作用:11, 白萝卜性味辛甘:11, 肺凉心:11, 苡仁米:11, 小鼠肉瘤:11, 碎小块:11, 橄榄萝卜:11, 稀稠粥:10, 银耳锅:10, 气管炎:10, 玉竹汁液:10, 快克:10, 汤肴:9, 脾虚湿盛:8, 马兜铃:8 瓷杯:8, 治疗:8, 咯血:7, 白萝卜:7, 鼻涕:7, 肺片:7, 急支糖浆:7, 蛇胆川贝:6, 宣肺:6, 普通人流感:6, 肺血管破裂:6, 抑制率:6, 气道:5, 热感冒:5, 白猪:5, 炎症:5, 温热服:5, 流感病毒:5, 绿脓杆菌:5, 阴虚咳喘:4, 恶阴:4, 猪肉杏仁汤治:4, 防尘口罩:4, 豆腐

[0133]

	<p>糖:4, 康尧:4, 熟羊脂:4, 喉癌:3, 山药杏仁粥:3, 痰量:3, 内加白糖:3, 痉挛性:3, 弱虚:3, 黄精冰糖:3, 黄河利亚:3, 细支气管炎:3, 麻醉性强效:3, 抵抗力:3 盖盘:3, 白痰:3, 散酒毒:3, 支气管粘膜:3, 乐信药业:3, 结核:3, 空腹开水:3, 生梨:3, 右美沙芬:3, 化痰药:3, 猪油蜜膏:3, 荞面蛋清:3, 开水碗:3, 组织胺:2, 机体损耗:2, 对症:2, 患儿:2, 患者:2, 微寒:2, 检查:2, 鲜良:2, 反射性:2, 理脾:2, 养阴清肺丸:2, 吸入:2, 含片:2, 排痰:2, 宰杀:2, 粳米:2, 大蒜水:2, 蜂蜜:2, 血痰:2, 甲型:2, 喉咙痒:2, 呼吸衰竭:2, 上海康尧:2, 沙氏鹿茸草:2, 枇杷:2, 中枢性因素:2, 马乳:2, 银耳:2, 保护性:2, 博列康尼:2, 病毒性感冒:2, 瑜伽洗鼻壶:2, 鼻咽处:2, 胸部胀满:2, 肺阴虚:2, 抗衰:2, 服法:2, 老年人正气:2, 交感神经药:1, 海石粉:1, 白开水:1, 糖水:1, 蛇胆陈皮:1, 麦冬:1, 肾咳喘:1, 瓜蒌:1, 子宫颈癌:1, 虚汗多:1, 合剂:1, 低热:1, CT:1, 例男性:1, 发汗:1, 阴润肺:1, 大汤碗:1, 口罩:1, 白前:1, 食物:1, 配伍:1, 本方清肝火:1, 肺泡壁弹性组织:1, 痰涎:1, 上火:1, 浮沫:1, 无花果:1, 保护性作用:1, 大蒜:1, 旋复花:1, 解毒片:1, 多发性炎症:1, 呼吸道疾病:1, 萝卜水:1, 肺汤加:1, 大肠杆菌:1, 气管黏膜:1, 儿科:1, 绿色部份:1, 支原体感染:1, 肝阴亏虚:1, 过敏:1, 止咳露:1, 患者食道 24 小时 PH:1, 气痛:1, 鼠曲草:1, 咽部黏膜:1, 组织表面鼻涕:1, 中医顾铭浩验方:1, 休息:1, 标本虚实:1, 鼻窦炎:1, 调味:1, 传染源:1, 病因:1, 柚肉:1, 生姜红糖水:1, 神经末梢感受器:1, 晾温:1, 调理:1, 气管异物:1, 桑菊饮:1, 解毒:1, 呕吐:1, 贝母瓜蒌散:1, 淡豆豉:1, 鲜竹沥口服液:1, 小青龙:1, 白稀痰:1, 本方益肾:1, 病菌:1,</p>
--	---

[0134]



		鲜藕:1, 咸易:1, 胸膜疾病:1, 气虚:1, 散寒:1, 头孢类:1 耳鼻喉科医生:1, 果菜汁:1, 纱布包:1, 生姜切片:1, 鸡屎 藤:1, 主治:1, 补肾:1, 鼻粘膜表面:1
鼻涕	医疗健 康	鼻涕:108, 鼻窦炎:29, 脑脊梁:19, 快克:17, 热感冒:15, 感冒药:15, 鼻漏:15, 鼻窦:14, 鼻腔粘膜:12, 洗鼻器:6, 鼻腔膜:5, 瑜伽洗鼻壶:5, 鼻咽处:5, 解毒片:4, 药物性 鼻炎:4 病毒性感冒:4, 治疗:4, 鼻中隔:3, 咽喉:3, 脓鼻 涕:3, 组织表面鼻涕:3, 感冒颗粒:3, 粘液性:3, 冷空气:3, 口罩:2, 多为上颌窦:2, 普通人流感:2, 急性鼻窦炎:2, 咽炎:2, 抵抗力:2, 鼻腔分泌物:2, 多可少:2 代茶频:2, 味麻滴鼻液:2, 季节性:2, 清毒作用:2, 维 C 银翘片:2, 鼻窦部位:2, 低温等离子:2, 鼻腔间歇性:2, 类型鼻炎:2, 鼻粘膜表面:2, 小儿感冒颗粒:1, 对症:1, 寿司芥末酱:1, 新康泰克:1, 痒感:1, 氨酚黄那敏颗粒:1, 姐姐症状差:1, RHINITIS:1, 水样:1, 吸入:1, 开水:1, 脱敏:1, 鼻腔粘 液腺:1, 热伤风:1, 千柏:1, 上火:1, 息斯敏:1, 鼻腔黏膜:1, 病毒大本营:1, 流感症状:1, 鼻腔分泌物涂片:1, 支气管 炎:1, 清热:1, 净鼻腔:1, 生理因素:1, 口腔病菌:1, 刺激 物:1, 考特:1, 发作:1, 甲型:1, 儿童感冒药效果:1, 多鼻 涕:1, 季节性鼻炎:1, 流脓:1, 全身病因:1, 通窍:1, 着凉:1, 宝宝:1, 病菌:1, 快克感冒药:1, 老干丝瓜:1, 生姜:1, 苏 理肺:1, 臭鼻症:1, 阵发性喷嚏:1

[0135] 第四步,根据待分析答案词语集合中的待分析答案词语,在第三步所选择得到的问答知识记录的基础上筛选出其包括的答案词语与待分析答案词语匹配的问答知识记录,进而得到所筛选出的问答知识记录的语义相关度。经分析可知,本例中与问答知识记录中的答案词语匹配的待分析答案词语包括:<口服,咳喘,小儿感冒颗粒,检查,止咳,治疗,流感症状,感冒颗粒>;

[0136] 再计算上述待分析的问答对的相关联程度可以得出,该待分析的问答对的相关联程度达到了 0.9 (在相关联程度取值范围为 0 ~ 1 的条件下);

[0137] 根据相关联程度获取所述待分析问答对的搜索排名。本例仅以一个待分析问答对

的相关联程度为例,在搜索结果包括多个问答对的情况下,可以对所述问答对从语义方面分别计算相关联程度,进而优化问答对的搜索排名,从而使相关联程度高的搜索结果排名靠前。

[0138] 需要说明的是:

[0139] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0140] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0141] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0142] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0143] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中所述的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如,在下面的权利要求书中,所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0144] 本发明的各个部件实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的优化问答对的搜索排名的装置中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如,计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上,或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0145] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

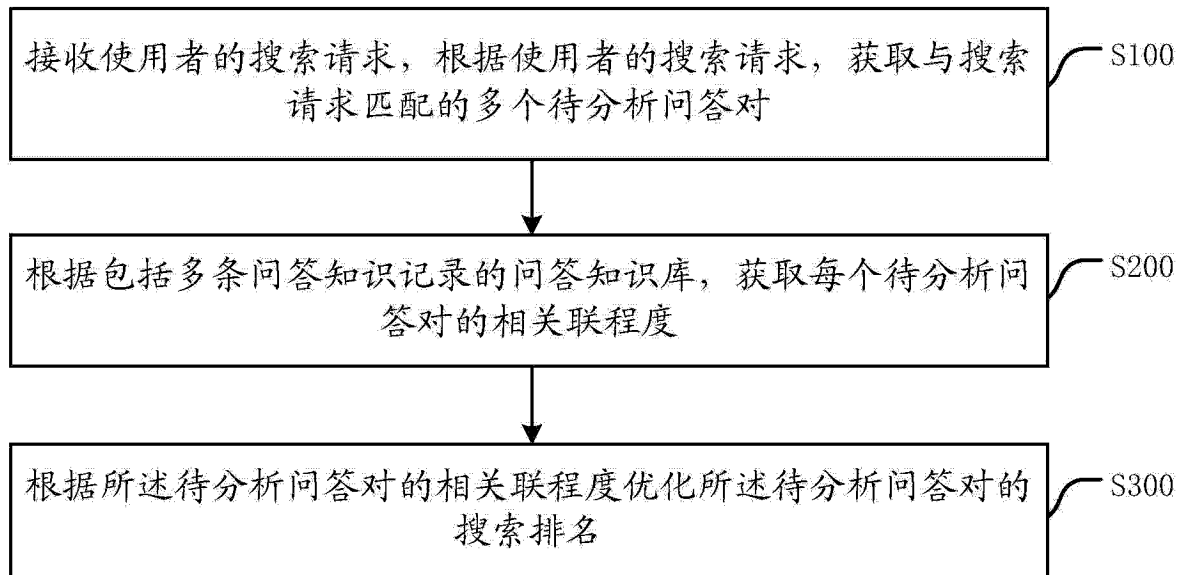


图 1

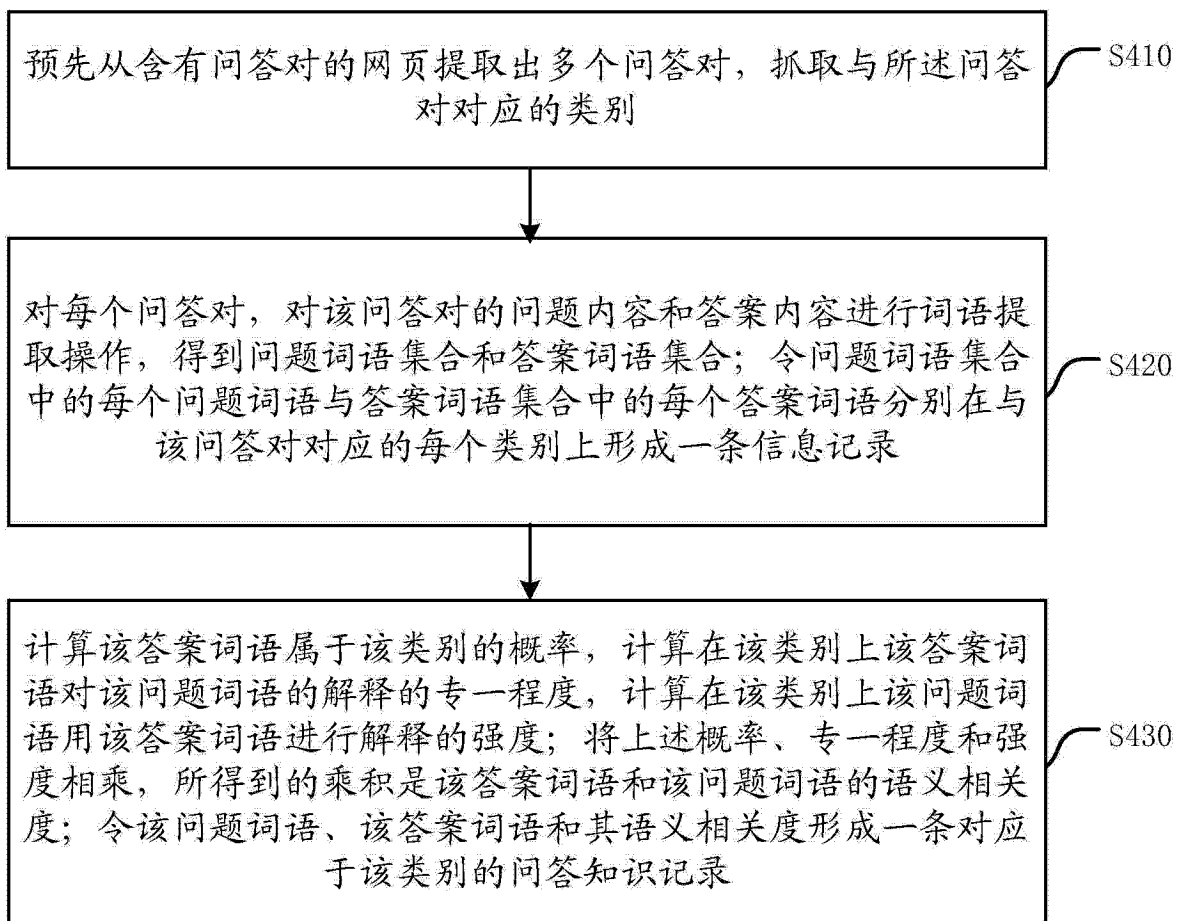


图 2

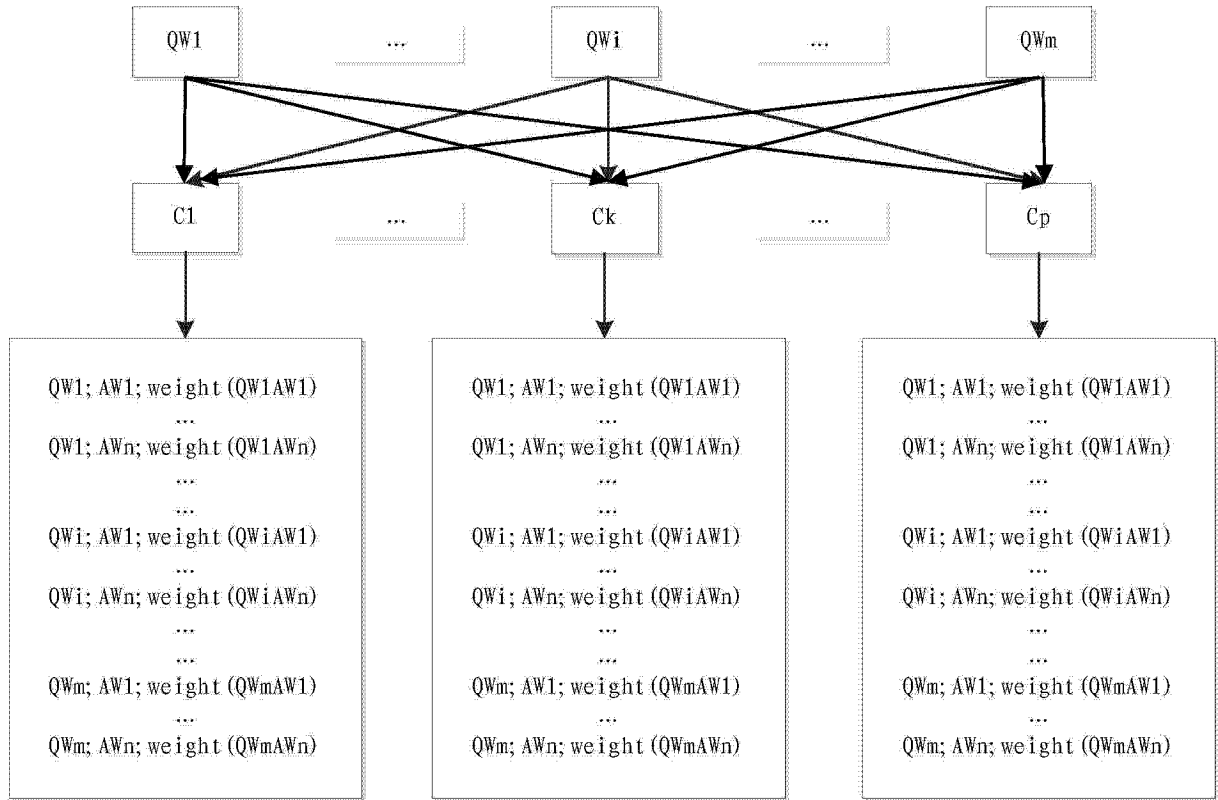


图 3

图1的步骤S200

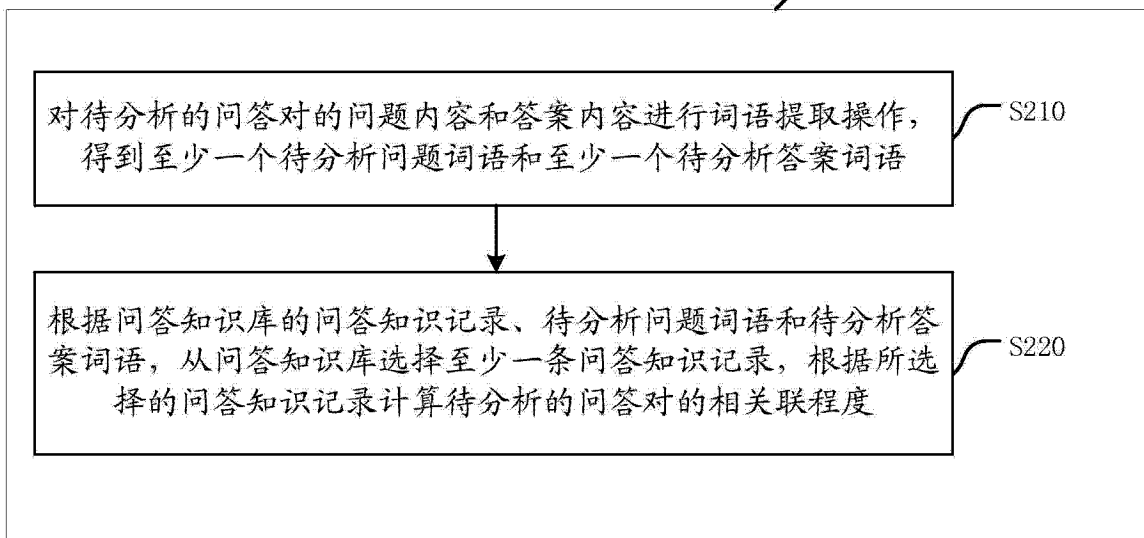


图 4

图4的步骤S220

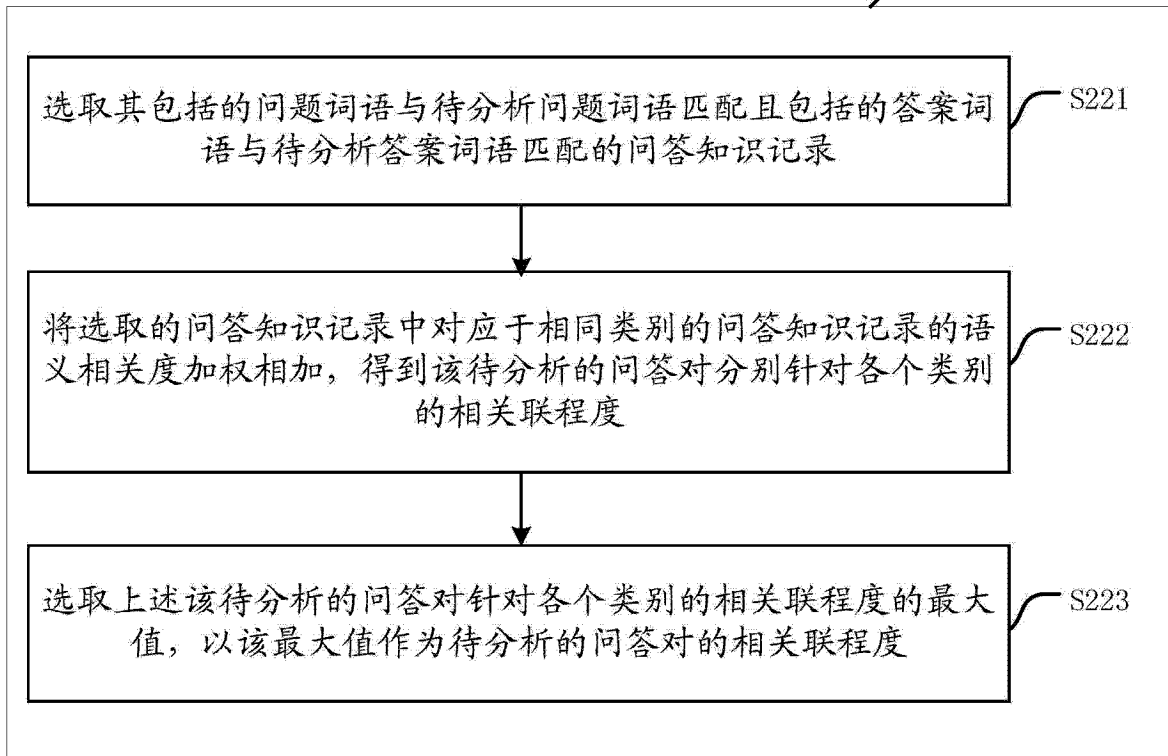


图5

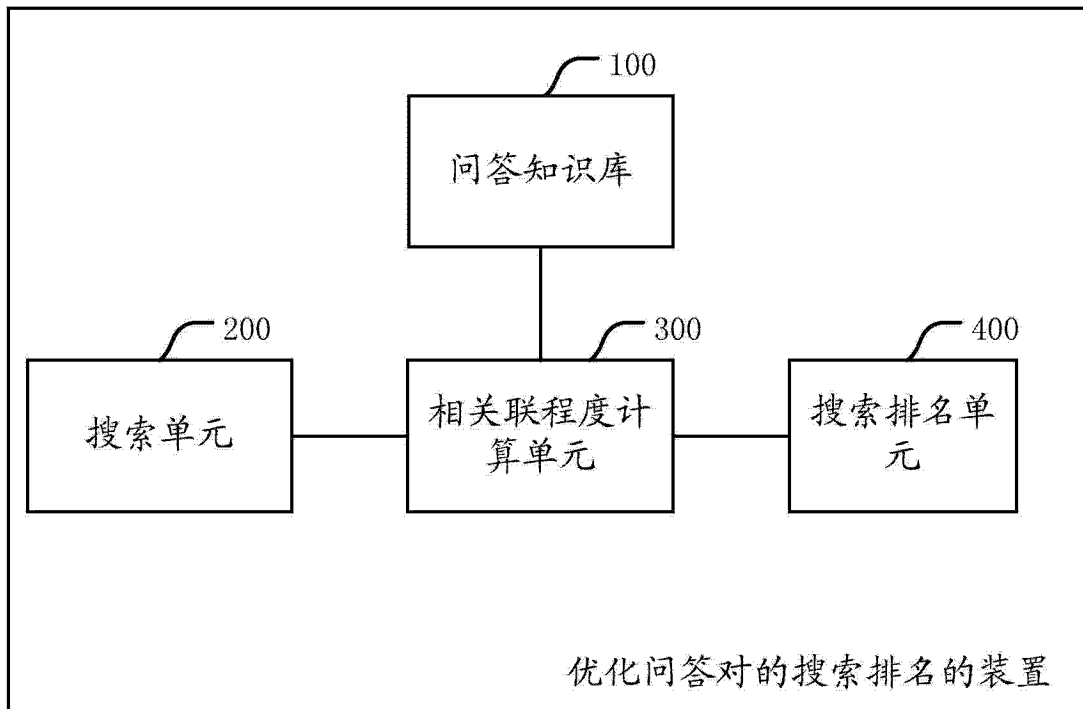


图6

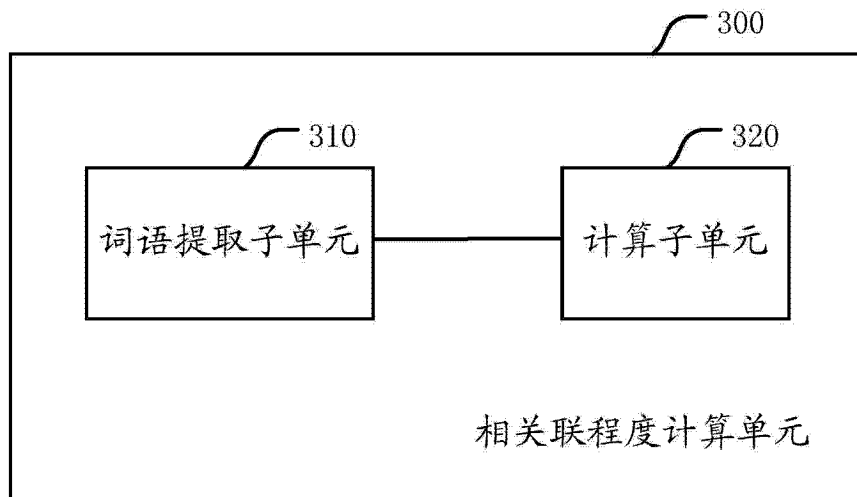


图 7

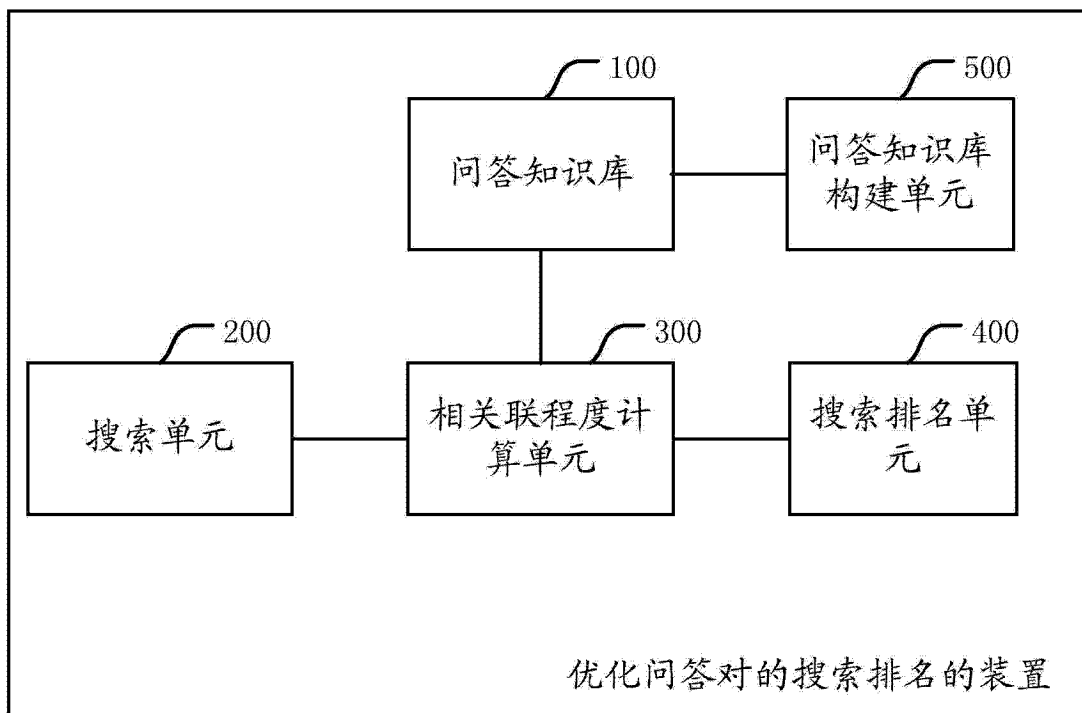


图 8