



(12) 发明专利

(10) 授权公告号 CN 101339551 B

(45) 授权公告日 2013. 01. 30

(21) 申请号 200710127431. 4

JP 特开 2006-106970 A, 2006. 04. 20, 全文.

(22) 申请日 2007. 07. 05

JP 特开 2005-63185 A, 2005. 03. 10, 全文.

(73) 专利权人 日电(中国)有限公司

审查员 徐长波

地址 北京市东城区东四十条甲 22 号南新仓
国际大厦 B 栋 12 层 1222 室 100007

(72) 发明人 丰强泽 齐红威 福岛俊一
梁邦勇 孟昕 刘宇光

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021

代理人 罗松梅

(51) Int. Cl.

G06F 17/30(2006. 01)

(56) 对比文件

CN 1794240 A, 2006. 06. 28, 说明书第 1 页第
11 行到第 3 页第 3 行, 第 4 页第 2 行到第 8 页第
13 行, 权利要求书.

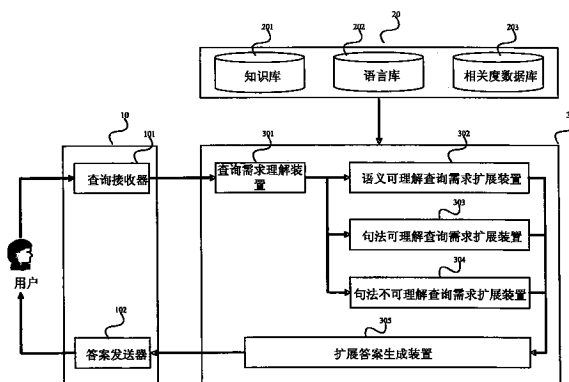
权利要求书 8 页 说明书 15 页 附图 21 页

(54) 发明名称

自然语言查询需求扩展设备及其方法

(57) 摘要

本发明提供了一种查询需求扩展设备, 包括:
查询需求理解装置, 用于根据用户的查询请求生成
显式查询需求; 以及查询需求扩展装置, 用于生
成与显式查询需求相关的隐式查询需求。其中查
询需求理解装置通过查找知识库和语言库生成
包括查询概念和问题类型的显式查询需求, 以
及查询需求扩展装置通过查找知识库, 语言库
和相关度数据库生成包括查询概念和问题类型
的隐式查询需求。本发明还提供了一种查询需求
扩展方法。基于本发明的查询需求扩展设备和
方法, 可以方便用户查询, 并向用户提供准确、
全面的查询答案。



1. 一种查询需求扩展设备,包括:

查询需求理解装置,用于根据用户的查询请求生成显式查询需求;以及
查询需求扩展装置,用于生成与显式查询需求相关的隐式查询需求,其中
查询需求理解装置通过查找知识库和语言库生成包括查询概念和问题类型的显式查询需求;以及

查询需求扩展装置通过查找知识库、语言库和相关度数据库生成包括查询概念和问题类型的隐式查询需求。

2. 如权利要求 1 所述的查询需求扩展设备,其中

查询需求理解装置通过对接收到的用户查询进行语义和句法上的分析来生成显式查询需求;以及

查询需求扩展装置通过查找位于所述设备内部或外部的需求相关数据库生成与显式查询需求相关的隐式查询需求。

3. 如权利要求 1 所述的查询需求扩展设备,其中查询需求理解装置包括:

分词单元,用于根据知识库和语言库,对用户查询进行分词;

匹配单元,用于将分词的用户查询与语言库进行匹配,以找到匹配文法;

语义一致性判断单元,用于根据知识库和语言库,判断用户查询中的概念是否有语义错误;

显式查询需求生成单元,用于根据匹配文法,生成与其对应的、包括查询概念和问题类型的显式查询需求;

需求类型判断单元,用于根据知识库和语言库,判断显式查询需求属于下列三种类型中的哪一种:语义可理解、句法可理解或句法不可理解。

4. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求。

5. 如权利要求 4 所述的查询需求扩展设备,其中语义可理解查询需求扩展装置包括:

相关问题类型获取单元,用于通过查找相关度数据库获取与显式查询需求的问题类型间服务相关度高的问题类型,作为检索到的相关问题类型集合;

相关查询概念获取单元,用于查找相关度数据库,获取与显式查询需求的查询概念相关度高的概念,作为检索到的相关查询概念集合;

相关查询需求筛选单元,用于将相关查询概念集合和相关问题类型集合相互组合得到相关查询需求集合,并查找知识库和语言库,去除语义错误的相关查询需求;以及

查询需求扩展生成单元,用于计算生成的相关查询需求与显式查询需求之间的相关度,并选择相关度高的相关查询需求作为隐式查询需求。

6. 如权利要求 5 所述的查询需求扩展设备,其中

查询需求扩展生成单元分别计算相关查询概念和查询概念之间的第一相关度以及相关问题类型和问题类型之间的第二相关度,并对第一相关度和第二相关度求和,从而得到相关查询需求与显式查询需求之间的相关度,其中相关查询需求包括相关查询概念和相关问题类型,显式查询需求包括查询概念和问题类型。

7. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求。

8. 如权利要求 7 所述的查询需求扩展设备,其中句法可理解查询需求扩展装置包括:

概念约束获取单元,用于根据显式查询需求所包括的问题类型,检索语言库,提取与问题类型对应的文法中定义的概念约束作为约束条件;

相关查询概念获取单元,用于利用相关度数据库获取与显式查询需求中所包括的查询概念之间相关度高的概念,并基于从概念约束获取单元获取的约束条件查找知识库,以选择出满足约束条件的查询概念;以及

查询需求扩展生成单元,用于将相关查询概念与显式查询需求中的问题类型组合生成隐式查询需求。

9. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

句法不可理解查询需求扩展装置,用于利用语言库,检索与显式查询需求中包含的问题类型相似的问题类型,和/或利用相关度数据库和知识库,查找与显式查询需求中包含的未识别查询概念相似的查询概念,从而获取包括相似查询概念和/或相似问题类型的隐式查询需求。

10. 如权利要求 9 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置包括:

相似问题类型获取单元,用于通过检索语言库,获取与显式查询需求中包含的问题类型相似的相似问题类型;

相似查询概念获取单元,用于根据相关度数据库中的字符串相似数据获取与未识别的查询概念之间字符串相似的查询概念,作为相似查询概念;以及

查询需求扩展生成单元,用于组合相似查询概念和相似问题类型,并生成隐式查询需求。

11. 如权利要求 10 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置还包括:

与相似问题类型获取单元连接的概念含义猜测单元,用于通过知识库和与相似问题类型对应的文法中定义的概念约束,猜测未识别查询概念所属的类别;以及

其中相似查询概念获取单元用于从具有和未识别查询概念相似的字符串的概念中选择属于同一猜测类别的概念作为相似概念。

12. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求;以及

句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类

型组合作为隐式查询需求。

13. 如权利要求 12 所述的查询需求扩展设备,其中语义可理解查询需求扩展装置包括:

相关问题类型获取单元,用于通过查找相关度数据库获取与显式查询需求的问题类型间服务相关度高的问题类型,作为检索到的相关问题类型集合;

相关查询概念获取单元,用于查找相关度数据库,获取与显式查询需求的查询概念相关度高的概念,作为检索到的相关查询概念集合;

相关查询需求筛选单元,用于将相关查询概念集合和相关问题类型集合相互组合得到相关查询需求集合,并查找知识库和语言库,去除语义错误的相关查询需求;以及

查询需求扩展生成单元,用于计算生成的相关查询需求与显式查询需求之间的相关度,并选择相关度高的相关查询需求作为隐式查询需求。

14. 如权利要求 13 所述的查询需求扩展设备,其中

查询需求扩展生成单元分别计算相关查询概念和查询概念之间的第一相关度以及相关问题类型和问题类型之间的第二相关度,并对第一相关度和第二相关度求和,从而得到相关查询需求与显式查询需求之间的相关度,其中相关查询需求包括相关查询概念和相关问题类型,显式查询需求包括查询概念和问题类型。

15. 如权利要求 12 所述的查询需求扩展设备,其中句法可理解查询需求扩展装置包括:

概念约束获取单元,用于根据显式查询需求所包括的问题类型,检索语言库,提取与问题类型对应的文法中定义的概念约束作为约束条件;

相关查询概念获取单元,用于利用相关度数据库获取与显式查询需求中所包括的查询概念之间相关度高的概念,并基于从概念约束获取单元获取的约束条件查找知识库,以选择出满足约束条件的相关查询概念;以及

查询需求扩展生成单元,用于将相关查询概念与显式查询需求中的问题类型组合生成隐式查询需求。

16. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求;以及

句法不可理解查询需求扩展装置,用于利用语言库,检索与显式查询需求中包含的问题类型相似的问题类型,和/或利用相关度数据库和知识库,查找与显式查询需求中包含的未识别查询概念相似的查询概念,从而获取包括相似查询概念和/或相似问题类型的隐式查询需求。

17. 如权利要求 16 所述的查询需求扩展设备,其中语义可理解查询需求扩展装置包括:

相关问题类型获取单元,用于通过查找相关度数据库获取与显式查询需求的问题类型间服务相关度高的问题类型,作为检索到的相关问题类型集合;

相关查询概念获取单元,用于查找相关度数据库,获取与显式查询需求的查询概念相关度高的概念,作为检索到的相关查询概念集合;

相关查询需求筛选单元,用于将相关查询概念集合和相关问题类型集合相互组合得到相关查询需求集合,并查找知识库和语言库,去除语义错误的相关查询需求;以及

查询需求扩展生成单元,用于计算生成的相关查询需求与显式查询需求之间的相关度,并选择相关度高的相关查询需求作为隐式查询需求。

18. 如权利要求 17 所述的查询需求扩展设备,其中

查询需求扩展生成单元分别计算相关查询概念和查询概念之间的第一相关度以及相关问题类型和问题类型之间的第二相关度,并对第一相关度和第二相关度求和,从而得到相关查询需求与显式查询需求之间的相关度,其中相关查询需求包括相关查询概念和相关问题类型,显式查询需求包括查询概念和问题类型。

19. 如权利要求 16 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置包括:

相似问题类型获取单元,用于通过检索语言库,获取与显式查询需求中包含的问题类型相似的相似问题类型;

相似查询概念获取单元,用于根据相关度数据库中的字符串相似数据获取与未识别的查询概念之间字符串相似的查询概念,作为相似查询概念;以及

查询需求扩展生成单元,用于组合相似查询概念和相似问题类型,并生成隐式查询需求。

20. 如权利要求 19 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置还包括:

与相似问题类型获取单元连接的概念含义猜测单元,用于通过知识库和与相似问题类型对应的文法中定义的概念约束,猜测未识别查询概念所属的类别;以及

其中相似查询概念获取单元用于从具有和未识别查询概念相似的字符串的概念中选择属于同一猜测类别的概念作为相似概念。

21. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求;以及

句法不可理解查询需求扩展装置,用于利用语言库,检索与显式查询需求中包含的问题类型相似的问题类型,和/或利用相关度数据库和知识库,查找与显式查询需求中包含的未识别查询概念相似的查询概念,从而获取包括相似查询概念和/或相似问题类型的隐式查询需求。

22. 如权利要求 21 所述的查询需求扩展设备,其中句法可理解查询需求扩展装置包括:

概念约束获取单元,用于根据显式查询需求所包括的问题类型,检索语言库,提取由文法所定义的约束条件;

相关查询概念获取单元,用于利用相关度数据库获取与显式查询需求中所包括的查询概念之间相关度高的概念,并基于从概念约束获取单元获取的约束条件查找知识库,以选择出满足约束条件的查询概念;以及

查询需求扩展生成单元,用于将相关查询概念与显式查询需求中的问题类型组合生成隐式查询需求。

23. 如权利要求 21 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置包括:

相似问题类型获取单元,用于通过检索语言库,获取与显式查询需求中包含的问题类型相似的相似问题类型;

相似查询概念获取单元,用于根据相关度数据库中的字符串相似数据获取与未识别的查询概念之间字符串相似的查询概念,作为相似查询概念;以及

查询需求扩展生成单元,用于组合相似查询概念和相似问题类型,并生成隐式查询需求。

24. 如权利要求 23 所述的查询需求扩展设备,其中句法不可理解查询需求扩展装置还包括:

与相似问题类型获取单元连接的概念含义猜测单元,用于通过知识库和与相似问题类型对应的文法中定义的概念约束,猜测未识别查询概念所属的类别;以及

其中相似查询概念获取单元用于从具有和未识别查询概念相似的字符串的概念中选择属于同一猜测类别的概念作为相似概念。

25. 如权利要求 1 所述的查询需求扩展设备,其中查询需求扩展装置包括:

语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求;

句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求;以及

句法不可理解查询需求扩展装置,用于利用语言库,检索与显式查询需求中包含的问题类型相似的问题类型,和/或利用相关度数据库和知识库,查找与显式查询需求中包含的未识别查询概念相似的查询概念,从而获取了包括相似查询概念和/或相似问题类型的隐式查询需求。

26. 如权利要求 1 所述的查询需求扩展设备,还包括:

扩展答案生成装置,用于利用隐式查询需求来检索语言库、知识库及一信息库,从而生成扩展查询答案。

27. 如权利要求 1 所述的查询需求扩展设备,还包括:

直接答案生成装置,用于利用显式查询需求来检索语言库、知识库及信息库,从而生成直接的查询答案。

28. 一种用于扩展查询需求的方法,包括:

查询需求理解步骤,通过查询需求理解装置根据用户的查询请求生成显式查询需求;以及

查询需求扩展步骤,通过查询需求扩展装置生成与显式查询需求相关的隐式查询需求,其中

查询需求理解步骤包括通过查找知识库和语言库生成包括查询概念和问题类型的显式查询需求的步骤；以及

查询需求扩展步骤包括通过查找知识库、语言库和相关度数据库生成包括查询概念和问题类型的隐式查询需求的步骤。

29. 如权利要求 28 所述的扩展查询需求的方法，其中

查询需求理解步骤包括通过对接收到的用户查询进行语义和句法上的分析来生成显式查询需求的步骤；以及

查询需求扩展步骤包括通过查找需求相关数据库生成与显式查询需求相关的隐式查询需求的步骤。

30. 如权利要求 28 所述的扩展查询需求的方法，其中查询需求理解步骤包括：

分词步骤，根据知识库和语言库，对用户查询进行分词；

匹配步骤，将分词的用户查询与语言库进行匹配，以找到匹配文法；

语义一致性判断步骤，根据知识库和语言库，判断用户查询中的概念是否有语义错误；

显式查询需求生成步骤，根据匹配文法，生成与其对应的、包括查询概念和问题类型的显式查询需求；

需求类型判断步骤，根据知识库和语言库，判断显式查询需求属于下列三种类型中的哪一种：语义可理解、句法可理解或句法不可理解。

31. 如权利要求 28 所述的扩展查询需求的方法，其中查询需求扩展步骤包括下列步骤中至少之一：

语义可理解查询需求扩展步骤，根据语义上可理解的显式查询需求检索相关度数据库，获得与语义上可理解的查询需求相关的相关查询需求集合，其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求；

句法可理解查询需求扩展步骤，基于知识库和相关度数据库，形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合，并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求；以及

句法不可理解查询需求扩展步骤，利用语言库，检索与显式查询需求中包含的问题类型相似的问题类型，和 / 或利用相关度数据库和知识库，查找与显式查询需求中包含的未识别查询概念相似的查询概念，从而获取包括相似查询概念和 / 或相似问题类型的隐式查询需求。

32. 如权利要求 31 所述的扩展查询需求的方法，其中语义可理解查询需求扩展步骤包括：

相关问题类型获取步骤，通过查找相关度数据库获取与显式查询需求的问题类型间服务相关度高的问题类型，作为检索到的相关问题类型集合；

相关查询概念获取步骤，查找相关度数据库，获取与显式查询需求的查询概念相关度高的概念，作为检索到的相关查询概念集合；

相关查询需求筛选步骤，将相关查询概念集合和相关问题类型集合相互组合得到相关查询需求集合，并查找知识库和语言库，去除语义错误的相关查询需求；以及

查询需求扩展生成步骤,计算生成的相关查询需求与显式查询需求之间的相关度,并选择相关度高的相关查询需求作为隐式查询需求。

33. 如权利要求 32 所述的扩展查询需求的方法,其中

查询需求扩展生成步骤包括分别计算相关查询概念和查询概念之间的第一相关度以及相关问题类型和问题类型之间的第二相关度,并对第一相关度和第二相关度求和的步骤,从而得到相关查询需求与显式查询需求之间的相关度,其中相关查询需求包括相关查询概念和相关问题类型,显式查询需求包括查询概念和问题类型。

34. 如权利要求 31 所述的扩展查询需求的方法,其中句法可理解查询需求扩展步骤包括:

概念约束获取步骤,根据显式查询需求所包括的问题类型,检索语言库,提取与问题类型对应的文法中定义的概念约束作为约束条件;

相关查询概念获取步骤,利用相关度数据库获取与显式查询需求中所包括的查询概念之间相关度高的概念,并基于从概念约束获取步骤获取的约束条件查找知识库,以选择出满足约束条件的查询概念;以及

查询需求扩展生成步骤,将相关查询概念与显式查询需求中的问题类型组合生成隐式查询需求。

35. 如权利要求 31 所述的扩展查询需求的方法,其中句法不可理解查询需求扩展步骤包括:

相似问题类型获取步骤,通过检索语言库,获取与显式查询需求中包含的问题类型相似的相似问题类型;

相似查询概念获取步骤,根据相关度数据库中的字符串相似数据获取与未识别的查询概念之间字符串相似的查询概念,作为相似查询概念;以及

查询需求扩展生成步骤,组合相似查询概念和相似问题类型,并生成隐式查询需求。

36. 如权利要求 35 所述的扩展查询需求的方法,其中句法不可理解查询需求扩展步骤还包括:

概念含义猜测步骤,通过知识库和与相似问题类型对应的文法中定义的概念约束,猜测未识别查询概念所属的类别;以及

其中相似查询概念获取步骤包括从具有和未识别查询概念相似的字符串的概念中选择属于同一猜测类别的概念作为相似概念的步骤。

37. 如权利要求 29 所述的扩展查询需求的方法,其中相关度数据库是由下列步骤中的至少之一生成的:

基于知识库,计算知识库中的概念之间的语义相关度,以便生成语义相关数据;

通过统计语料库,计算知识库中的概念之间的互信息,以便生成统计相关数据;

根据词典,求出其中各词语之间的字符串相似度,以便生成字符串相似数据;

基于用户查询日志,求出各查询所匹配的文法对应的问题类型之间的相关度,以便生成服务相关数据。

38. 如权利要求 29 所述的扩展查询需求的方法,其中知识库是由下列步骤生成的:

根据各领域信息库,抽取其中的概念、属性及概念间的关系,从而生成各领域知识库;

根据同义词典、简称规则库、多语言词典和电子地图,抽取同义映射关系、语言映射关

系和空间映射关系,从而生成映射知识库。

39. 如权利要求 29 所述的扩展查询需求的方法,其中语言库是由下列步骤生成的:

收集各领域中的用户查询句子,计算句子间的相似度并依据相似度对句子进行聚类,生成各领域查询语言;

计算各领域查询语言间的相似度,抽取公用查询语言。

40. 如权利要求 29 所述的扩展查询需求的方法,还包括:

扩展答案生成步骤,利用隐式查询需求来检索语言库、知识库及一信息库,从而生成扩展查询答案。

41. 如权利要求 29 所述的扩展查询需求的方法,还包括:

直接答案生成步骤,利用显式查询需求来检索语言库、知识库及信息库,从而生成直接的查询答案。

自然语言查询需求扩展设备及其方法

技术领域

[0001] 本发明涉及信息查询领域,更具体地,涉及一种自然语言查询需求扩展设备以及一种自然语言查询需求扩展方法。

背景技术

[0002] 目前存在多种根据用户发送的查询请求提供查询结果的系统。但是当用户进行自然语言查询时,通常期望知道比他/她所发送的查询请求的更多的信息。例如,当一个用户查询海龙在哪里时,那么除了海龙的地址外,可能用户也想知道怎么到达海龙。所以,根据用户的当前查询可以推测用户潜在的查询需求,即,隐式查询需求。但是现有的查询系统仅能够根据用户的当前查询,生成与之相关的直接的查询答案,而没有考虑用户的潜在查询需求,以对用户的当前查询进行扩展。

[0003] 已知存在一些相关的查询系统及方法。US20050289124 提出一种自然语言查询处理的系统和方法,它可以处理用户的英文自然语言查询,生成相应的语义标注,并从知识库中检索到对应的信息。

[0004] US6411950 提出一种关键词查询的扩展方法,它利用查询日志,将关键词查询中的抽象词扩展为查询频度较高的具体词。如将查询词“car”扩展为(“car audio”OR “rent a car”OR “car price”OR...)。

[0005] US6480843 也提出一种关键词查询的扩展方法,它利用词语间的语义相似度,将关键词查询中的词语进行扩展。如将关键词查询“car dealer”扩展为[(“car”OR “automobile”OR “auto”OR “Ford”)]AND(“Dealer”OR “Showroom”OR “SalesOffice”)。

[0006] 综合现有的查询系统或方法,可以发现其或者只能分析出显式查询需求,如US20050289124。或者只能对关键词查询中的词语进行扩展,如US6411950,US6480843。目前还没有方法可以对自然语言查询的当前显式查询需求查询请求进行有效扩展。

发明内容

[0007] 为了解决上述问题,提出了本发明。本发明提出了一种自然语言查询需求扩展设备和方法,可以对接收到的用户的自然语言查询进行分析,获得隐式查询需求,从而扩展了用户的查询需求并生成扩展的答案信息。通过扩展用户的查询需求,可以帮助自然语言查询需求扩展设备更好地理解查询用户的意图,发掘用户的潜在兴趣与需求,使用户得到充分、全面的答案信息。

[0008] 根据本发明第一方面,提供了一种查询需求扩展设备,包括:

[0009] 查询需求理解装置,用于根据用户的查询请求生成显式查询需求;以及

[0010] 查询需求扩展装置,用于生成与显式查询需求相关的隐式查询需求。

[0011] 根据本发明第二方面,提供了一种查询需求扩展设备,其中查询需求理解装置通过对接收到的用户查询进行语义和句法上的分析来生成显式查询需求;以及

[0012] 查询需求扩展装置通过查找位于所述设备内部或外部的需求相关数据库生成与

显式查询需求相关的隐式查询需求。

[0013] 根据本发明第三方面,提供了一种查询需求扩展设备,其中查询需求理解装置通过查找知识库和语言库生成包括查询概念和问题类型的显式查询需求;以及

[0014] 查询需求扩展装置通过查找知识库,语言库和相关度数据库生成包括查询概念和问题类型的隐式查询需求。

[0015] 根据本发明第四方面,提供了一种查询需求扩展设备,其中查询需求理解装置包括:

[0016] 分词单元,用于根据知识库和语言库,对用户查询进行分词;

[0017] 匹配单元,用于将分词的用户查询与语言库进行匹配,以找到匹配文法;

[0018] 语义一致性判断单元,用于根据知识库和语言库,来判断用户查询中的概念是否有语义错误;

[0019] 显式查询需求生成单元,用于根据匹配文法,生成与其对应的、包括查询概念和问题类型的显式查询需求。

[0020] 需求类型判断单元,用于根据知识库和语言库,来判断显式查询需求属于下列三种类型中的哪一种:语义可理解、句法可理解或句法不可理解。

[0021] 根据本发明第五方面,提供了一种查询需求扩展设备,其中查询需求扩展装置包括:

[0022] 语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求。

[0023] 根据本发明第六方面,提供了一种查询需求扩展设备,其中查询需求扩展装置包括:

[0024] 句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求。

[0025] 根据本发明第七方面,提供了一种查询需求扩展设备,其中查询需求扩展装置包括:

[0026] 句法不可理解查询需求扩展装置,用于利用语言库,检索与显式查询需求中包含的问题类型相似的问题类型,和/或利用相关度数据库和知识库,查找与显式查询需求中包含的未识别查询概念相似的查询概念,从而获取包括相似查询概念和/或相似问题类型的隐式查询需求。

[0027] 根据本发明第八方面,提供了一种查询需求扩展设备,包括:

[0028] 语义可理解查询需求扩展装置,用于根据语义上可理解的显式查询需求检索相关度数据库,获得与语义上可理解的查询需求相关的相关查询需求集合,其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求;

[0029] 句法可理解查询需求扩展装置,用于基于知识库和相关度数据库,形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合,并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求。

题类型组合作为隐式查询需求 ; 以及

[0030] 句法不可理解查询需求扩展装置, 用于利用语言库, 检索与显式查询需求中包含的问题类型相似的问题类型, 和 / 或利用相关度数据库和知识库, 查找与显式查询需求中包含的未识别查询概念相似的查询概念, 从而获取了包括相似查询概念和 / 或相似问题类型的隐式查询需求。

[0031] 根据本发明第九方面, 提供了一种用于扩展查询需求的方法, 包括:

[0032] 查询需求理解步骤, 根据用户的查询请求生成显式查询需求 ; 以及

[0033] 查询需求扩展步骤, 生成与显式查询需求相关的隐式查询需求。

[0034] 根据本发明第十方面, 提供了一种用于扩展查询需求的方法, 其中

[0035] 查询需求理解步骤包括通过对接收到的用户查询进行语义和句法上的分析来生成显式查询需求的步骤 ; 以及

[0036] 查询需求扩展步骤包括通过查找需求相关数据库生成与显式查询需求相关的隐式查询需求的步骤。

[0037] 根据本发明第十一方面, 提供了一种用于扩展查询需求的方法, 其中查询需求理解步骤包括通过查找知识库和语言库生成包括查询概念和问题类型的显式查询需求的步骤 ; 以及

[0038] 查询需求扩展步骤包括通过查找知识库, 语言库和相关度数据库生成包括查询概念和问题类型的隐式查询需求的步骤。

[0039] 根据本发明第十二方面, 提供了一种用于扩展查询需求的方法, 其中查询需求扩展步骤包括下列步骤中至少之一:

[0040] 语义可理解查询需求扩展步骤, 根据语义上可理解的显式查询需求检索相关度数据库, 获得与语义上可理解的查询需求相关的相关查询需求集合, 其中包括包含了相关查询概念和相关问题类型以及查询概念和相关问题类型之一的隐式查询需求 ;

[0041] 句法可理解查询需求扩展步骤, 基于知识库和相关度数据库, 形成与句法上可理解的显式查询需求中的查询概念相关度高的相关查询概念集合, 并利用知识库从相关查询概念集合中选择出满足预定约束条件的至少一个相关查询概念与显式查询需求的问题类型组合作为隐式查询需求 ; 以及

[0042] 句法不可理解查询需求扩展步骤, 利用语言库, 检索与显式查询需求中包含的问题类型相似的问题类型, 和 / 或利用相关度数据库和知识库, 查找与显式查询需求中包含的未识别查询概念相似的查询概念, 从而获取包括相似查询概念和 / 或相似问题类型的隐式查询需求。

附图说明

[0043] 图 1a 是根据本发明的一个实施例的自然语言查询需求扩展设备的示意图 ;

[0044] 图 1b 是需求相关数据库的一个示例 ;

[0045] 图 1c 是根据本发明的一个实施例的自然语言查询需求扩展方法的流程图 ;

[0046] 图 2 是根据本发明的自然语言查询需求扩展设备的示意图 ;

[0047] 图 3(a) 是根据本发明的知识库的示意图 ;

[0048] 图 3(b) 是根据本发明的知识库的层次结构示意图 ;

- [0049] 图 4 是根据本发明的语言库的示意图；
- [0050] 图 5a 是根据本发明的相关度数据库的示意图；
- [0051] 图 5b 是根据本发明的相关度数据库生成装置的示意图；
- [0052] 图 5c 是相关度数据库生成装置生成相关度数据库的方法的流程图；
- [0053] 图 6a 是根据本发明的查询需求理解装置的示意图；
- [0054] 图 6b 是根据本发明的查询需求理解方法的流程图；
- [0055] 图 6c 是根据本发明进行查询需求理解的一个示例；
- [0056] 图 7a 是根据本发明的语义可理解查询需求扩展装置的示意图；
- [0057] 图 7b 是根据本发明的语义可理解查询需求扩展方法的流程图；
- [0058] 图 8a 是根据本发明的句法可理解查询需求扩展装置的示意图；
- [0059] 图 8b 是根据本发明的句法可理解查询需求扩展方法的流程图；
- [0060] 图 9a 是根据本发明的句法不可理解查询需求扩展装置的示意图；
- [0061] 图 9b 是根据本发明的句法不可理解查询需求扩展方法的流程图；
- [0062] 图 10a-10c 是三个示例；
- [0063] 图 11 是根据本发明另一实施例的自然语言查询需求扩展设备的示意图。

具体实施方式

[0064] 下面,将参考附图描述本发明的优选实施例。在附图中,相同的元件将由相同的参考符号或数字表示。此外,在本发明的下列描述中,将省略对已知功能和配置的具体描述,以避免使本发明的主题不清楚。

[0065] 图 1a 示出根据本发明的一个实施例的自然语言查询需求扩展设备。根据用户输入的查询请求,该自然语言查询需求扩展设备通过检索需求相关数据库 5 生成隐式查询需求并生成相应的扩展答案。该自然语言查询需求扩展设备包括用于接收用户的查询请求的查询接收器 1;查询需求理解装置 2;检索位于自然语言查询需求扩展设备内部或者外部的需求相关数据库生成隐式查询需求的查询需求扩展装置 3 以及扩展答案生成装置 4。

[0066] 图 1c 示出了生成扩展答案的自然语言查询需求扩展方法的流程图。在步骤 1',查询接收器 1 接收来自用户的查询请求。在步骤 2',由查询需求理解装置 2 对显式查询需求查询请求进行语义和句法上的分析,找出查询请求中包括显式查询需求包含的查询概念和问题类型,作为用户的显式查询需求。查询需求理解装置 2 可以根据启发式规则或机器学习的方法,来判定用户的显式查询需求所对应的问题类型,然后将用户查询请求中的名词作为查询概念。首先需要确定问题类型。目前主要有两种方法,一种方法是通过手工总结的启发式规则加以判定,另一种方法为问题自动分类。此外也可以采用本发明图 6c 所示的方法来获取显式查询需求。

[0067] 1) 启发式规则方法。根据用户查询请求中出现的疑问词及其他词(动词、名词、副词、形容词等)来总结启发式规则。例如规则“在/位于,哪里/哪儿→地址”表示用户查询请求中有“在”或“位于”,而且有“哪里”或“哪儿”时,问题类型通常为地址。例如规则“如何/怎么,到/去/到达→路径”表示用户查询请求中有“如何”或“怎么”,而且有“到”或“去”或“到达”时,问题类型通常为路径。

[0068] 根据启发式规则,可以判定用户查询请求所对应的问题类型。

[0069] 2) 自动分类方法。利用机器学习的方法将一个问题识别到一个或多个语义类别中。主要方法有利用词作为特征项、利用句法信息以及语义信息。但自动分类得到的结果不够准确,目前的自动分类方法的准确率一般都小于 90%。

[0070] 在获取了问题类型之后,由查询需求理解装置 2 取出用户查询请求中剩下的名词作为查询概念。

[0071] 例如对于用户的查询请求“海龙大厦在哪里”,其包含“哪里”,则根据启发式规则“在 / 位于,哪里 / 哪儿→地址”,判定其问题类型为地址,并且抽出名词“海龙大厦”作为概念。因此获得显式查询需求为“海龙大厦 ;地址”。

[0072] 在步骤 3',根据获得的显式查询需求,首先由查询需求扩展装置 3 根据用户查询中的查询概念及问题类型,分别检索需求相关数据库 5,得到相关的查询概念集合与问题类型集合。图 1b 示出了需求相关数据库 5 的一个示例。该需求相关数据库 5 中存储了各个概念之间的相关度,以及问题类型之间的相关度。之后,查询需求扩展装置 3 将上面得到的相关查询概念与相关问题类型进行组合,生成相关查询需求集合,并计算每个相关查询需求与当前查询需求间的相关度。计算公式为:需求“ $c_1 ; t_1$ ”和“ $c_2 ; t_2$ ”间的相关度 = c_1 和 c_2 的相关度 * t_1 和 t_2 的相关度,其中, c_1 和 c_2 是概念, t_1 和 t_2 是问题类型。

[0073] 可以按需求相关度来递减排序相关查询需求,将相关度大于某阈值的相关查询需求作为最终的隐式查询需求。

[0074] 例如对于显式查询需求“太平洋大厦 ;地址”,其所有相关查询需求为“海龙大厦 ;地址”(相关度 0.85),“太平洋 ;地址”(相关度 0.6),“太平洋大厦 ;路径”(相关度 0.9),“海龙大厦 ;路径”(相关度 0.765),“太平洋 ;路径”(相关度 0.54)。若阈值取 0.8,则最终的隐式查询需求为“太平洋大厦 ;路径”和“海龙大厦 ;地址”。

[0075] 最后在步骤 4',根据隐式查询需求,来生成扩展答案。

[0076] 可以由多种方法生成需求相关数据库 5,(1) 利用现有的词相似度计算方法。基于 HowNet、同义词林等现有的语义词典,计算词间的相似度。(2) 采用本发明图 5c 所示的生成方法。计算语义相关、统计相关、字符串相似、服务相关,最后取平均值作为概念之间或问题类型之间的相关度。例如“海龙大厦”和“中关村大街”的语义相关度为 1,统计相关度为 0.7,则二者的相关度为 $(1+0.7)/2 = 0.85$ 。

[0077] 图 2 示出自然语言查询需求扩展设备的结构图。自然语言查询需求扩展设备具有包括查询接收器 101 和答案发送器 102 的接收 / 发送装置 10,存储器 20 以及查询分析扩展装置 30。在自然语言查询需求扩展设备通过查询接收器 101 接收到用户的自然语言查询请求时,将该自然语言查询请求传送到查询分析扩展装置 30 进行分析扩展处理,得到隐式查询需求,并生成扩展的查询需求答案。之后将处理后的查询需求扩展答案经过答案发送器 102 发送给用户。存储器存储有知识库 201,语言库 202 以及相关度数据库 203。查询分析扩展装置 30 包括:用于判断用户的自然语言查询请求是语义可理解查询需求、句法可理解查询需求还是句法不可理解查询需求的查询需求理解装置 301,用于生成隐式查询需求的语义可理解查询需求扩展装置 302、句法可理解查询需求扩展装置 303、以及句法不可理解查询需求扩展装置 304 以及利用隐式查询需求来检索语言库、知识库以及信息数据库从而生成扩展的答案的扩展答案生成装置 305。查询分析扩展装置 30 利用知识库 201,语言库 202,相关度数据库 203 来对自然语言查询进行扩展处理。

[0078] 可以理解,虽然图 2 示出了知识库 201,语言库 202 以及相关度数据库 203 和信息库(未示出)设置在自然语言查询需求扩展设备的内部,但是,可选的,知识库,语言库、相关度数据库以及信息库也可以设置在自然语言查询需求扩展设备的外部。此外,自然语言查询需求扩展设备还可以包括直接答案生成装置,可以根据用户的显式查询需求直接生成相应的查询答案。从而,利用本发明的自然语言查询需求扩展设备不但可以获得扩展的查询答案,而且可以获得用户的直接的查询答案。

[0079] 存储器 20 存储有知识库 201,语言库 202,相关度数据库 203 以及信息数据库。其中,知识库 201 包括用于处理用户查询的语义知识。语言库 202 包括用于理解用户查询的语言模型。相关度数据库 203 包括用于扩展用户查询需求的相关性数据。信息数据库包括用于生成扩展答案和直接答案的具体领域数据。查询分析扩展装置 30 基于知识库、语言库以及相关度数据库可以从用户的自然语言查询中获得显式查询需求,并分析出用户的扩展查询需求,从而向用户反馈更充分、更完全的答案信息。下面将参考图 3(a),图 3(b),图 4,图 5a 以及图 5b 来分别对知识库 201,语言库 202 以及相关度数据库 203 进行描述。之后再对如何利用这些数据库进行自然语言查询需求扩展进行详细描述。

[0080] 1. 知识库

[0081] 图 3(a) 示出知识库的一个示例结构。知识库由一组领域知识库和一个映射知识库组成。每个领域对应于一个领域知识库,用来记录该领域的语义知识。例如地图领域对应于地图知识库,黄页领域对应于黄页知识库。不同领域的知识并非完全孤立,它们经常存在许多关联,由一个映射知识库来记录不同领域知识库的概念间的各种关联。

[0082] 一般的,领域知识库包括三部分:概念部分、属性部分和关系部分。

[0083] 1) 概念部分。概念部分包括该领域内所有的类别和个体。

[0084] 例如在地图知识库中,概念部分包含电子地图中示出的所有地理类别和个体;类别用于描述各地点所属的类别,例如电子大厦、路都是类别,个体用于描述电子地图上的一个具体地点,例如,海龙大厦、中关村大街都是个体。

[0085] 2) 属性部分。属性定义了概念的性质。例如,经度、纬度等。每个属性都有一个项‘类型’表明属性的取值类型。

[0086] 3) 关系部分。关系部分描述了概念间的各种关联,如 isa 用来描述类别间、个体与类别间的隶属关系,near 用来描述个体间的空间相邻关系。每个关系都有两个项“类型”和“性质”。“类型”表明关系的取值类型,“性质”主要有两种:传递性、对称性。若 $R(x, y) \& R(y, z) \rightarrow R(x, z)$, 则 R 满足传递性,例如 isa 关系满足传递性;若 $R(x, y) \rightarrow R(y, x)$, 则 R 满足对称性,例如 near 关系满足对称性。

[0087] 例如在地图知识库中,“isa(海龙大厦,电子大厦)”表示海龙大厦是一个电子大厦;再如“near(海龙大厦,中关村大街)”表示中关村大街在海龙大厦附近。

[0088] 映射知识库只包含关系部分,具体包括同义映射关系部分、语言映射关系部分和空间映射关系部分。

[0089] 1) 同义映射关系部分。在同义词或缩略语之间进行映射。例如“synonymous(海龙大厦,海龙电子大厦)”表示二者是同义词。

[0090] 2) 语言映射关系部分。在同一概念的不同语言翻译之间进行映射。例如“chinese-english(中关村大街,Zhongguancun Street)”表示后者是前者的英文翻译。

[0091] 3) 空间映射关系部分。在空间相关的概念之间进行映射。例如“near(海龙大厦, 320 路公交车)”表示二者距离很近。

[0092] 图 3(b) 示出知识库的一个层次结构图, 其中, 该层次结构是以各领域知识库的概念部分中的概念为结点, 以各领域知识库的关系部分中的对应关系以及映射知识库中的对应关系为弧而自动生成的。

[0093] 知识库的生成分为两个步骤:

[0094] 首先根据服务提供者提供的信息库, 自动生成各领域知识库。其方法为: 1) 利用现有的信息提取算法, 从信息库中抽取各个概念, 以形成包括类别和个体的概念部分; 2) 利用现有的信息提取算法, 从信息库中提取数据字段做为属性部分; 3) 利用现有的信息提取算法, 从信息库和语义词典中抽取概念之间的关系以形成包括类别间关系、个体间关系以及类别与个体间关系的关系部分。例如对地图领域, 为了生成地图知识库, 首先从电子地图数据库中抽取各个概念, 以形成概念部分, 其中概念包括类别和个体。之后, 从电子地图数据库中抽取概念之间的隶属关系以形成关系部分, 其中关系包括类别之间、个体与类别之间以及个体与个体之间的隶属关系和空间关系。可以基于现有的 GIS 平台所提供的功能进行上述处理。

[0095] 其次根据各领域知识库中的概念, 自动生成映射知识库。其方法为:

[0096] 1) 根据同义词典和简称规则库生成同义映射关系部分。同义词典包括概念及相关同义词, 例如“银谷大厦, 银谷电子大厦”表示“银谷大厦”和“银谷电子大厦”是同义词。简称规则库包括各短语的缩略语, 例如“abbreviate(北京大学, 北大)”表示“北京大学的简称是北大”。根据简称规则可以判断两个概念是否同义, 例如根据 abbreviate(北京大学, 北大)”和“abbreviate(附属小学, 附小)”, 可以得到 synonymous(北京大学附属小学, 北大附小)。

[0097] 2) 根据多语言词典生成语言映射关系部分。

[0098] 3) 根据电子地图的 GIS 函数生成空间映射关系部分。

[0099] 2. 语言库

[0100] 语言库包括用于理解用户查询的语言模型。语言库由领域查询语言和公用查询语言组成。每个领域对应一个领域查询语言, 例如路况领域对应于路况领域查询语言, 公交领域对应于公交领域查询语言, 黄页领域对应于黄页领域查询语言。公用查询语言总结了各领域查询语言的共同的查询特征, 它可以被各领域查询语言进行参数继承。领域查询语言和公用查询语言的表示方法是一样的。

[0101] 图 4 示出了语言库的一个示例。查询语言包括四部分: 问题类型、继承、文法和动作。

[0102] ①问题类型部分。问题类型可以为任意字符串。例如“路况”。

[0103] ②继承部分。继承部分描述了该问题类型可以继承的文法, 例如“路况: 属性值查询(? cate = 路 | 桥, ? A = 路况)”表示“路况”的文法可以通过对参数? cate 和? A 的赋值来继承问题类型“属性值查询”的文法, 即“<? C(路 | 桥)>[的]<路况>”也属于“路况”的文法。

[0104] ③文法部分。文法部分总结了关于该问题类型的所有可能的自然语言查询, 例如“<? C(路 | 桥)><堵 | 拥堵 | 畅通>”。文法部分由一组文法词组成, 例如“? C”, “堵”,

“拥堵”，“畅通”。语法部分中定义了一些特殊的符号：

[0105] ● “|”指的是“或”逻辑运算。

[0106] ● $\langle ? X(\text{cons}_1 | \dots | \text{cons}_m) \rangle$ 指 X 是一概念， $\text{cons}_i (i = 1 \dots m)$ 是一个约束条件，用于约束概念 X 的取值范围， cons_i 一般是知识库中的一个类别，此处 X 只能被知识库中属于任意一个类别 cons_i 的个体来代替，即 X 需要满足 $\text{isa}(X, \text{cons}_i)$ 。例如， $\langle ? C(\text{路} | \text{桥}) \rangle$ 只能被知识库中的路或桥的个体（如中关村大街）来代替。

[0107] ● “[]”指的是在“[”和“]”中间的内容可有可无。

[0108] ④动作部分。动作部分定义了该问题类型对应的查询动作，用于指明“当用户查询符合该问题类型的定义时，生成什么样的查询动作”，例如“路况查询（？ C）”。查询动作提供对用户查询的理解结果，可以由答案生成装置来执行。

[0109] 语言库的生成分为两个步骤：

[0110] 首先，针对每个领域，收集各领域的用户可能的查询句子，并利用现有的句子相似度计算方法，计算各查询句子间的相似度，按相似度聚类，生成领域查询语言的语法部分。其次对每种语法，人工定义其问题类型和动作部分。例如根据“路况领域”中的查询句子“联想桥堵吗”，“北四环拥堵吗”，“学院路畅通吗”可得到问题类型“路况”的语法“ $\langle ? C(\text{路} | \text{桥}) \rangle \langle \text{堵} | \text{拥堵} | \text{畅通} \rangle$ ”（“吗”是语气词不计入）。

[0111] 其次，计算各领域查询语言的语法部分的相似度，抽取公用查询语言，并生成相应领域查询语言的继承部分。例如根据“路况”语法“ $\langle ? C(\text{路} | \text{桥}) \rangle [\text{的}] \langle \text{路况} \rangle$ ”和“地址”语法“ $\langle ? C(\text{地点}) \rangle [\text{的}] \langle \text{地址} \rangle$ ”，可得到“属性值查询”的语法“ $\langle ? C(? \text{cate}) \rangle [\langle \text{的} \rangle] \langle ? A \rangle$ ”，而“路况”语法的继承部分则为“属性值查询（？ cate = 路 | 桥，？ A = 路况）”。

[0112] 3. 相关度数据库

[0113] 图 5a 示出了相关度数据库的一个示例。相关度数据库中记录了所有对象（知识库中的概念、语言库中的问题类型）之间的相关性的相关度数据。主要考虑四种相关性：

[0114] a. 语义相关。表示不同概念之间的语义相关性。例如概念“海龙大厦”与“中关村大街”都是地点，且有空间邻近关系，则二者语义相关。具有语义相关性的两个概念既可以是个体，也可以是类别。

[0115] b. 统计相关。表示不同概念在统计意义上的相关性。例如“海龙大厦”与“中关村大街”在文本中经常一起出现，则二者统计相关。

[0116] c. 字符串相似。表示不同词语在描述形式上的相似性。例如“太平洋”与“太平洋大厦”在词语描述形式上相关，“太平洋”是“太平洋大厦”的子串，而“海龙”是“海龙大厦”的子串，所以，“太平洋”与“太平洋大厦”，“海龙”与“海龙大厦”字符串相似。

[0117] d. 服务相关。表示不同问题类型所提供的服务的相关性。例如“地址”和“路径”都是关于位置服务的，所以“地址”和“路径”服务相关。

[0118] 图 5b 示出了相关度数据库生成装置的示意结构图。该相关度数据库生成装置包括：利用知识库来生成语义相关性数据的语义相关度计算装置 42，利用语料库生成统计相关性数据的统计相关度计算装置 44，根据通用词典生成字符串相似性数据的字符串相似度计算装置 46 以及分析用户查询日志，生成服务相关性数据的服务相关度计算装置 48。

[0119] 图 5c 是相关度数据库生成装置生成相关度数据库的方法的流程。

[0120] 在 S401, 语义相关度计算装置 42 基于知识库计算不同概念之间的语义相关度。参考图 3(b), 图 3(b) 中示出了以概念为结点, 关系为弧形成的知识库的结构层次图。而不同概念之间的语义相关度可以由它们之间的最短路径来衡量, 概念之间的最短路径越短, 说明语义越相关。

[0121] 语义相关度的计算公式如下:

[0122]

$$\text{语义相关度} = \frac{1}{\text{Len}(\text{MinPath}(x, y))} \quad (1)$$

[0123] 其中 $\text{Len}(\text{MinPath}(x, y))$ 为 x 与 y 之间的最短路径长度, 即从 x 到 y 之间的最少弧数。例如图 3(b) 中“海龙大厦”与“中关村大街”间的最短路径为“海龙大厦 -Near- 中关村大街”, 它们之间只有一条弧, 故最短路径长度为 1, 因此“海龙大厦”和“中关村大街”的语义相关度为 1。

[0124] 在 S402, 统计相关度计算装置 44 计算统计相关度。通过对语料库进行统计来计算两个概念的互信息。本发明所使用的语料库包含从 web 上收集来的大约 2,000,000 个网页。互信息是信息论中的测度, 用来衡量两个事件的关联程度。概念间的互信息越高, 说明统计越相关。统计相关度的计算公式如下:

[0125]

$$\text{统计相关度} = \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (2)$$

[0126] 其中 $P(x, y)$ 为 x 和 y 共同出现的概率, $P(x)$ 和 $P(y)$ 分别为 x 和 y 独立出现的概率。例如“海龙大厦”与“中关村”共同出现的概率为 0.98, 分别出现的概率为 0.14, 则“海龙大厦”与“中关村”的统计相关度为 0.7。

[0127] 在 S403, 字符串相似度计算装置 46 计算字符串相似度。本发明使用一个公知的通用词典, 其中存放了所有常见的词语 (包括知识库中的所有概念), 然后开始计算通用词典中各词间的字符串相似度。可以采用基于公知的编辑距离方法来求出字符串的相似度, 即, 要使两个概念串变成一样, 需要多少次字符的替换、插入或删除操作。操作次数越少, 字符串相似度越高。

[0128] 字符串相似度的计算公式如下:

[0129]

$$\text{字符串相似度} = 1 - \frac{D(x, y)}{\text{Max}(\text{Len}(x), \text{Len}(y))} \quad (3)$$

[0130] 其中 $D(x, y)$ 为使 x 与 y 相同的最少操作次数, $\text{Max}(\text{Len}(x), \text{Len}(y))$ 为 x 和 y 的字符串长度的最大值。例如在“太平洋”结尾插入两个字符, 即可变成“太平洋大厦”, 因此“太平洋”和“太平洋大厦”的字符串相似度为 $1 - 2/5 = 0.6$ 。

[0131] 在 S404, 服务相关度计算装置 48 计算服务相关度。已经记录了所有用户的查询日志, 每个用户的查询日志包括该用户提交过的所有查询及对应的查询需求。服务相关度计算装置 48 对所有用户的查询日志进行分析, 求出各查询的问题类型被用户连续询问的频率。两个问题类型越常被用户连续询问, 则二者服务越相关。服务相关度计算公式如下:

[0132] 服务相关度 = $P(Q_{n+1} \text{ 的问题类型为 } y | Q_n \text{ 的问题类型为 } x)$

[0133] (4)

[0134] 上式表示用户上次查询的问题类型是 x 时, 下次查询的问题类型是 y 的概率。例

如用户在问完某地点的地址后,经常会接着问如何到达该地点,因此,说明“地址”问题类型和“路径”问题类型的服务相关度很高,经计算得到“地址”和“路径”的服务相关度为 0.9。

[0135] 图 6a 示出查询需求理解装置的结构图。查询需求理解装置 301 包括分词单元 3011,与分词单元连接的匹配单元 3012,与匹配单元连接的语义一致性判断单元 3013,与语义一致性判断单元连接的显式查询需求生成单元 3014 以及与显式查询需求生成单元连接的需求类型判断单元 3015。分词单元 3011 通过查找知识库 201 中各领域知识库的概念部分、语言库 202 中各领域查询语言和公用查询语言的语法部分,对用户查询进行分词。匹配单元 3012 用于将分词后的用户查询,与语言库进行匹配。语义一致性判断单元 3013 用于基于知识库 201 中各领域知识库的关系部分,判断用户查询中的概念是否满足问题类型对应的语法中定义的概念约束。显式查询需求生成单元 3014 根据匹配语法,获得与之对应的问题类型,并将查询概念与问题类型组合形成显式查询需求。需求类型判断单元 3015 用于判断用户查询所属的需求类型:语义可理解、句法可理解还是句法不可理解。

[0136] 图 6b 示出查询需求理解装置执行查询需求理解的流程图。在 S50,分词单元 3011 通过查找知识库 201 中各领域知识库的概念部分、语言库 202 中各领域查询语言和公用查询语言的语法部分,对用户查询进行分词。

[0137] 例如用户查询“海龙大厦堵吗?”的分词结果为“海龙大厦;堵”,其中“海龙大厦”是地图知识库中的个体,“堵”是路况领域查询语言中的语法词。

[0138] 在 S52,匹配单元 3012 将分词后的用户查询,与语言库进行匹配,以找到与用户查询相匹配的语法。匹配顺序为:先匹配公用查询语言,若有匹配语法,则再参数匹配可继承该语法的领域查询语言,若没有可匹配的公用查询语言,则直接匹配各领域查询语言。

[0139] 例如将用户查询“海龙大厦的路况”(其分词结果为“海龙大厦;的;路况”)与语言库匹配的过程为:先匹配公用查询语言,其匹配结果为“属性值查询:<? C(? cate)><[<的><? A>”];再继续匹配继承“属性值查询”的领域查询语言,其匹配结果为“路况:属性值查询(? cate = 路 | 桥, ? A = 路况)”。

[0140] 再如将用户查询“海龙大厦堵吗?”与语言库匹配的结果为“路况:<? C(路 | 桥)><堵 | 拥堵 | 畅通>”。

[0141] 在 S54,语义一致性判断单元 3013 基于知识库 201 中各领域知识库的关系部分,判断用户查询中的概念是否满足所匹配语法中定义的概念约束。约束判断方法为根据 isa 关系判断用户查询中的概念 C 是否满足语法中定义的概念约束“ $cons_1 | \dots | cons_m$ ”之一,即是否有 $isa(C, cons_i)$,其中 $i = 1 \dots m$,若不满足则用户查询存在语义错误(例如,可以使用图 3(b) 所示的层次关系图来帮助进行判断)。例如对于用户查询“海龙大厦堵吗?”,查询概念为“海龙大厦”,匹配语法中的相应概念约束为“路 | 桥”,但由于不存在 $isa(\text{海龙大厦}, \text{路})$ 和 $isa(\text{海龙大厦}, \text{桥})$,因此语义一致性判断单元判断出用户查询中存在语义错误。

[0142] 在 S56,显式查询需求生成单元 3014 根据匹配语法得到查询概念及相应的问题类型,并生成显式查询需求。显式查询需求由查询概念及问题类型组成。参考图 6c 的示例,从用户查询“海龙大厦堵吗?”中提取出显式查询需求“海龙大厦;路况”。

[0143] 在 S58,需求类型判断单元 3015 根据显式查询需求,利用下述方法来判断该显式查询需求属于下列三种类型中的哪一种:语义可理解、句法可理解或句法不可理解。

[0144] ①判断查询概念是否认识。查找知识库,看各查询概念是否在知识库中某领域知

识库的概念部分中。若显式查询需求中没有查询概念,则认为存在不认识的查询概念。

[0145] ②判断问题类型是否认识。查找语言库,看是否在其中。若显式查询需求中的问题类型为空,则认为存在不认识的问题类型。

[0146] ③判断语义一致性。先根据问题类型,查找语言库,找到问题类型对应的文法中定义的概念约束;然后基于知识库,判断显式查询需求中的概念是否满足上述概念约束。约束判断方法与语义一致性判断单元 3013 中的约束判断方法相同。

[0147] ④需求类型判断。根据以上模块的执行情况,判断用户查询所属的类型。

[0148] ●若需求类型判断单元认识所有概念和问题类型,且无语义错误,则为语义可理解查询;

[0149] ●若需求类型判断单元认识所有概念和问题类型,但存在语义错误,则认为是句法可理解查询;

[0150] ●若需求类型判断单元有不认识的概念和 / 或问题类型,则认为是句法不理解查询。

[0151] 对于图 6c 的示例的用户查询“海龙大厦堵吗?”,需求类型判断单元 3015 判断其类型为句法可理解查询。

[0152] 图 7a 是根据本发明的语义可理解查询需求扩展装置的一个示例结构图。语义可理解查询需求扩展装置利用用户语义上可理解的查询需求,获得其隐式查询需求。该语义可理解查询需求扩展装置包括:相关问题类型获取单元 3021,用于通过查找相关度数据库获取与用户的当前查询的问题类型间服务相关度高的问题类型,作为检索到的相关问题类型集合;相关查询概念获取单元 3022,用于在相关问题类型获取单元未获取到相关问题类型时,查找相关度数据库,获取与当前查询的概念的相关度高的概念,作为检索到的相关查询概念集合;相关查询需求筛选单元 3023,用于将相关查询概念集合和相关问题类型集合相互组合得到相关查询需求集合,并查找知识库和语言库,去除语义错误的相关查询需求;以及查询需求扩展生成单元 3024,用于计算生成的相关查询需求与用户的当前查询之间的相关度,并选择相关度高的相关查询需求作为用户的扩展查询。

[0153] 图 7b 示出了语义可理解查询的需求扩展过程。首先,在 S601,由相关问题类型获取单元 3021 获取相关问题类型。相关问题类型获取单元 3021 针对显式查询需求所包括的问题类型,检索相关度数据库中的服务相关性数据,获取与显式查询需求的问题类型之间的服务相关度高的问题类型集合,作为相关问题类型集合。

[0154] 在 S602,如果没有检索到较为相关的问题类型,则针对显式查询需求所述的查询概念,相关查询概念获取单元 3022 检索相关度数据库中的语义相关性数据和统计相关性数据,获取与显式查询需求所述查询概念之间相关度高的所有概念,作为相关查询概念集合。

[0155] 在 S603,相关查询需求筛选单元 3023 筛选出有用的相关查询需求。首先根据上面得到的相关查询概念集合和相关问题类型集合,将其两两组合得到相关查询需求集合,其中,相关查询需求由相关查询概念和相关问题类型组成。其次,相关查询需求筛选单元 3023 通过查找知识库和语言库,检查每个相关查询需求中的概念是否满足该相关查询需求中的问题类型在语言库中定义的概念约束,若不满足,则说明该相关查询需求有语义错误,需要将其删除。其中,相关查询需求筛选单元 3023 所采用的概念约束判断方法与查询需求理解

装置所采用的概念约束判断方法一样。

[0156] 最后,在 S604,查询需求扩展生成单元 3024 对于筛选得到的各个相关查询需求,计算其与用户显式查询需求之间的相关度,其值等于相关查询需求所述概念与显式查询需求所述概念间的相关度以及相关查询需求所述问题类型与显式查询需求所述问题类型间的相关度之和。之后,查询需求扩展生成单元 3024 根据相关度大小对相关查询需求进行排序,选取相关度高的相关查询需求,作为隐式查询需求。隐式查询需求可以是多个。

[0157] 例如对用户查询“海龙大厦在哪里?”,其显式查询需求为“海龙大厦 ;地址”,由于问题类型“地址”和“路径”之间的服务相关度很高,而且海龙大厦也符合路径的文法“< 如何 | 怎样 | 怎么 > 去 | 到 | 到达 > ? C(地点)>”中的概念约束“地点”。因此可生成扩展查询“海龙大厦 ;路径”。

[0158] 此外,相关问题类型获取单元 3021 也可以在 S601 之后,将获取的相关问题类型与显式查询需求所包括的概念组合,作为隐式查询需求。图 8a 示出了句法可理解查询需求扩展装置的示意图。句法可理解查询需求扩展装置基于用户的显式查询需求利用用户的句法可理解的查询,对其查询进行扩展。句法可理解查询需求扩展装置 303 包括通过检索语言库,获取显式查询需求的查询概念属于知识库中的何种类别的概念约束获取单元 3031,利用相关度数据库中的语义相关性数据和统计相关性数据,获得与显式查询需求的查询概念相关的概念的相关查询概念获取单元 3032 以及将相关查询概念与显式查询需求的问题类型组合生成隐式查询需求的查询需求扩展生成单元 3033。

[0159] 图 8b 示出了句法可理解查询需求扩展装置执行句法可理解查询需求扩展的流程图。

[0160] 在 S701,概念约束获取单元 3031 根据显式查询需求的问题类型,检索语言库,从而提取出该问题类型对应的文法中定义的概念约束,即用户查询概念应该隶属于知识库中的哪些类别。相关查询概念获取单元 3032 基于知识库和相关度数据库,获取与显式查询需求的查询概念相关的概念。相关查询概念获取单元 3032 根据显式查询需求的查询概念,检索相关度数据库中的语义相关性数据和统计相关性数据,找到与显式查询需求的查询概念之间相关度高的所有概念的集合。之后,相关查询概念获取单元 3032 基于概念约束获取单元得到的概念约束,通过查找知识库,筛选出满足约束的相关查询概念。相关查询概念获取单元所采用的概念约束判断方法与查询需求理解装置所采用的概念约束判断方法相同。

[0161] 在 S703,查询需求扩展生成单元 3033 根据获取的相关查询概念,与显式查询需求的匹配问题类型组合生成隐式查询需求。

[0162] 例如对用户查询“海龙大厦堵吗?”,其显式查询需求为“海龙大厦 ;路况”,其中查询概念“海龙大厦”不满足问题类型“路况”对应的文法中定义的概念约束“路 | 桥”,而与“海龙大厦”在语义和统计上都相关的概念“中关村大街”是一条路,符合约束“路 | 桥”,因此可生成隐式查询需求“中关村大街 ;路况”。

[0163] 图 9a 是根据本发明的句法不可理解查询需求扩展装置的示意图。句法不可理解查询需求扩展装置对于句法不理解的查询,根据用户的显式查询需求,利用相关度数据库中的字符串相似数据获得其隐式查询需求。句法不可理解查询需求扩展装置包括用于通过检索语言库,获取与用户查询相似的问题类型的相似问题类型获取单元 3041 ;用于对用户查询中未识别的概念,通过知识库和语言库,猜测该未识别概念的含意的概念含义猜测单

元 3042 ;根据相关度数据库中的字符串相似数据以及知识库,获取与未识别概念相似的概念的相似查询概念获取单元 3043 ;以及生成隐式查询需求的查询需求扩展生成单元 3044。

[0164] 图 9b 示出了句法不可理解查询需求扩展装置执行的句法不可理解查询需求扩展方法的流程图。如果查询需求理解装置 301 既不认识显式查询需求中的概念,也不认识显式查询需求中的问题类型,则在 S801,相似问题类型获取单元 3041 根据查询需求理解装置对用户查询的分词结果,检索语言库,以从中检索到与用户查询最相似的文法,并提取出与该文法对应的问题类型作为相似问题类型。

[0165] 在 S802,概念含义猜测单元 3042 对于用户查询中未识别的概念,利用基于上下文的含义以及基于文法约束的概念含义来猜测未识别概念所隶属的类别。其中,基于上下文的含义猜测包括:对于用户查询中出现的每个词,概念含义猜测单元 3042 在知识库中进行查找。若该词在知识库中出现,而且从知识库中找到了与该词有语义关系的类别,则可以猜测未识别概念与该类别之间可能也存在着某种语义关联。基于文法约束的概念含义猜测包括:根据由相似问题类型获取单元 3041 得到的相似问题类型,由概念含义猜测单元 3042 检索语言库,从而提取出该问题类型对应的文法中定义的概念约束,即用户查询概念应该隶属于知识库中的哪些类别,则未识别概念可能也隶属于这些类别。然后,概念含义猜测单元 3042 综合基于上下文的含义猜测结果以及基于文法约束的概念含义猜测结果,求其交集,从而得到未识概念应属的类别。

[0166] 在 S803,相似查询概念获取单元 3043 根据相关度数据库和知识库,获取与未识别概念相似的概念。首先,相似查询概念获取单元 3043 基于相关度数据库中的字符串相似数据,获取所有与未识别概念之间字符串相似度较高的概念;之后如果概念含义猜测单元猜测出了未识概念应属的类别,那么这里需要从所有字符串相似的概念中,筛选出符合步骤 S802 获得的类别的概念。约束判断方法与查询需求理解装置所采用的约束判断方法相同。

[0167] 在 S804,查询需求扩展生成单元 3044 根据上面得到的相似概念和相似问题类型,组合生成隐式查询需求。

[0168] 例如对用户查询“我想买台电脑,怎么去太平洋?”,系统不认识“太平洋”,也没有可匹配的查询问题类型。首先,找到与用户查询相似的文法“文法:<如何|怎样|怎么>去|到|到达><? C(地点)>”,并取出该文法对应的问题类型“路径”;其次,猜测“太平洋”的含义,从图 3(b) 所示的知识库中可以看出,“电子大厦”与用户查询中的词“电脑”有关系“sell(电子大厦,电脑)”,而且相似问题类型“路径”对应文法中的概念约束为“地点”,而且“isa(电子大厦,地点)”,因此可以猜测“太平洋”属于类别“电子大厦”;然后,获取相似概念,从图 6a 中可以发现“太平洋”与“太平洋大厦”字符串相似,并且从图 3(b) 中可以看出“太平洋大厦”也属于类别“电子大厦”,因此我们将“太平洋大厦”作为相似概念;最后,生成隐式查询需求“太平洋大厦;路径”。

[0169] 如果查询需求理解装置 301 不认识显式查询需求中的概念,但认识显式查询需求中的问题类型,则概念含义猜测单元 3042 对于用户查询中未识别的概念,利用基于上下文的含义以及基于文法约束的概念含义来猜测未识别概念所隶属的类别。相似查询概念获取单元 3043 基于相关度数据库中的字符串相似数据,获取所有与未识别概念之间字符串相似度较高的概念,作为相似概念。如果概念含义猜测单元 3042 猜测出未识别概念应属的类别,那么需要从所有字符串相似的概念中,筛选出符合所获得的类别的概念。约束判断

方法与查询需求理解装置所采用的约束判断方法相同。

[0170] 之后,由查询需求扩展生成单元 3044 将得到的相似概念和显式查询需求中的问题类型进行组合生成隐式查询需求。

[0171] 如果查询需求理解装置 301 不认识显式查询需求中的问题类型,但认识显式查询需求中的概念,则相似问题类型获取单元 3041 根据查询需求理解装置对用户查询的分词结果,检索语言库,以从中检索到与用户查询最相似的文法,并提取出与该文法对应的问题类型作为相似问题类型。之后,由查询需求扩展生成单元 3044 将得到的相似问题类型和显式查询需求中的概念进行组合生成隐式查询需求。

[0172] 在对用户的查询进行了语义可理解查询、句法可理解查询以及句法不可理解查询处理之后,扩展答案生成装置 305 生成与隐式查询需求对应的扩展答案。具体步骤包括:首先,根据隐式查询需求中的问题类型,检索语言库,得到该问题类型对应的查询动作,再将隐式查询需求中的查询概念替换到查询动作的相应位置,从而生成隐式查询需求所对应的具体动作;然后根据具体动作来检索知识库及后台信息库,从而得到相应的答案信息,具体实现方法取决于具体的服务领域和信息库的格式,这里可以使用公知的方法,具体不再赘述。例如,对于扩展查询需求“海龙大厦;路径”,其中路径问题类型的查询动作为“路径查询(? C)”,将“? C”替换为“海龙大厦”后,得到对应的具体动作“路径查询(海龙大厦)”,然后经检索得到答案“320 路公交车可以到海龙大厦”。

[0173] 在对用户的查询进行了查询需求理解之后,直接答案生成装置生成与显式查询需求直接对应的答案。具体步骤包括:若用户查询属于语义可理解查询,则根据显式查询需求检索语言库、知识库及后台信息库,从而生成相应答案,具体检索和生成方法与扩展答案生成的方法相同;若用户查询属于句法可理解查询或句法不可理解查询,则生成相应的错误提示信息。例如对于显式查询需求“海龙大厦,路况”,属于句法可理解查询,系统生成直接答案“海龙大厦不是路或桥,系统找不到它的路况信息”。

[0174] 图 10a-10c 分别示出了如何利用知识库、语言库以及相关度数据库进行语义可理解查询、句法可理解查询以及句法不可理解查询的三个示例。

[0175] 图 11 示出根据本发明另一实施例的自然语言查询需求扩展设备的示意图。与图 2 所示的自然语言查询需求扩展设备比较,图 11 示出的自然语言查询需求扩展设备不包括句法可理解查询需求扩展装置和句法不可查询需求扩展装置。出于清楚描述的目的,省略了对上述装置的类似的描述。此外,虽然图 11 示出了知识库 201,语言库 202 以及相关度数据库 203 设置在自然语言查询需求扩展设备的内部,但是,可选的,知识库可以设置在自然语言查询需求扩展设备的外部。以及,自然语言查询需求扩展设备可以仅包括语义可理解查询需求扩展装置、句法可理解查询需求扩展装置和句法不可查询需求扩展装置中的至少之一。

[0176] 根据本发明另一实施例,自然语言查询需求扩展设备也可以只包括句法可理解查询需求扩展装置或者句法不可理解查询需求扩展装置。

[0177] 以及,可以理解的,虽然以中文路况查询为例,示出了路况查询的扩展处理,但是发明的查询需求扩展设备还可以应用于信息检索、智能教学、推荐系统等。以及对于本领域技术人员显而易见的是本发明也可以用于其它语言的路况查询的扩展处理,例如,英语和日语等。

[0178] 尽管已经结合示例描述了本发明,本领域的技术人员可以理解的是,在不偏离本发明的原理和精神的前提下,可以在本实施例中进行改变,本发明的范围由所附权利要求及其等价物限定。

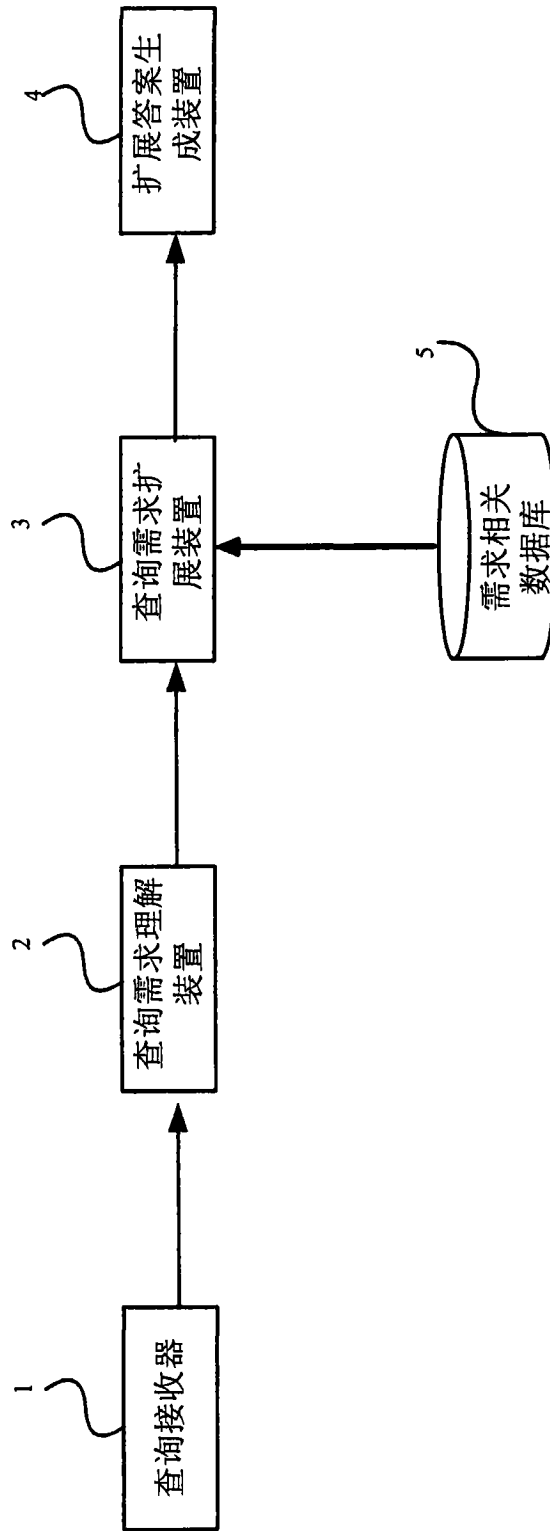


图 1a

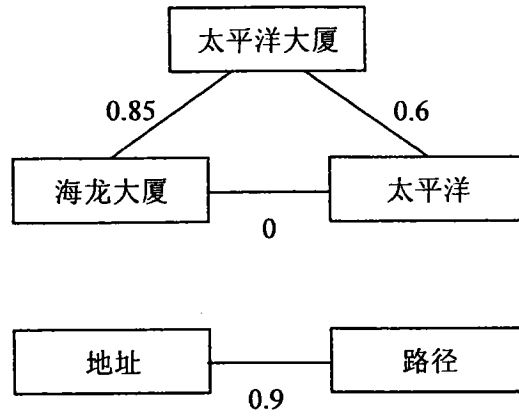


图 1b

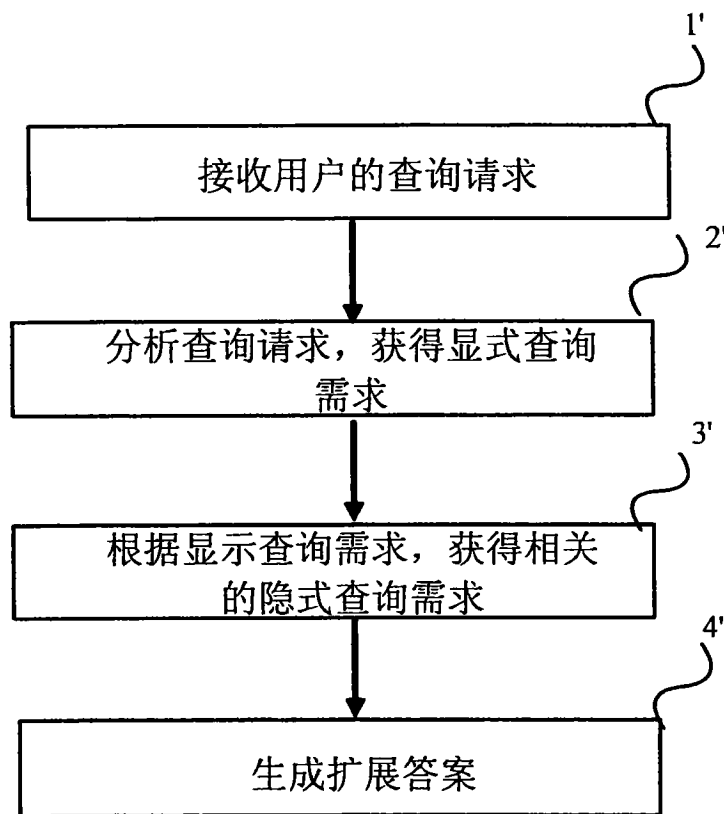


图 1c

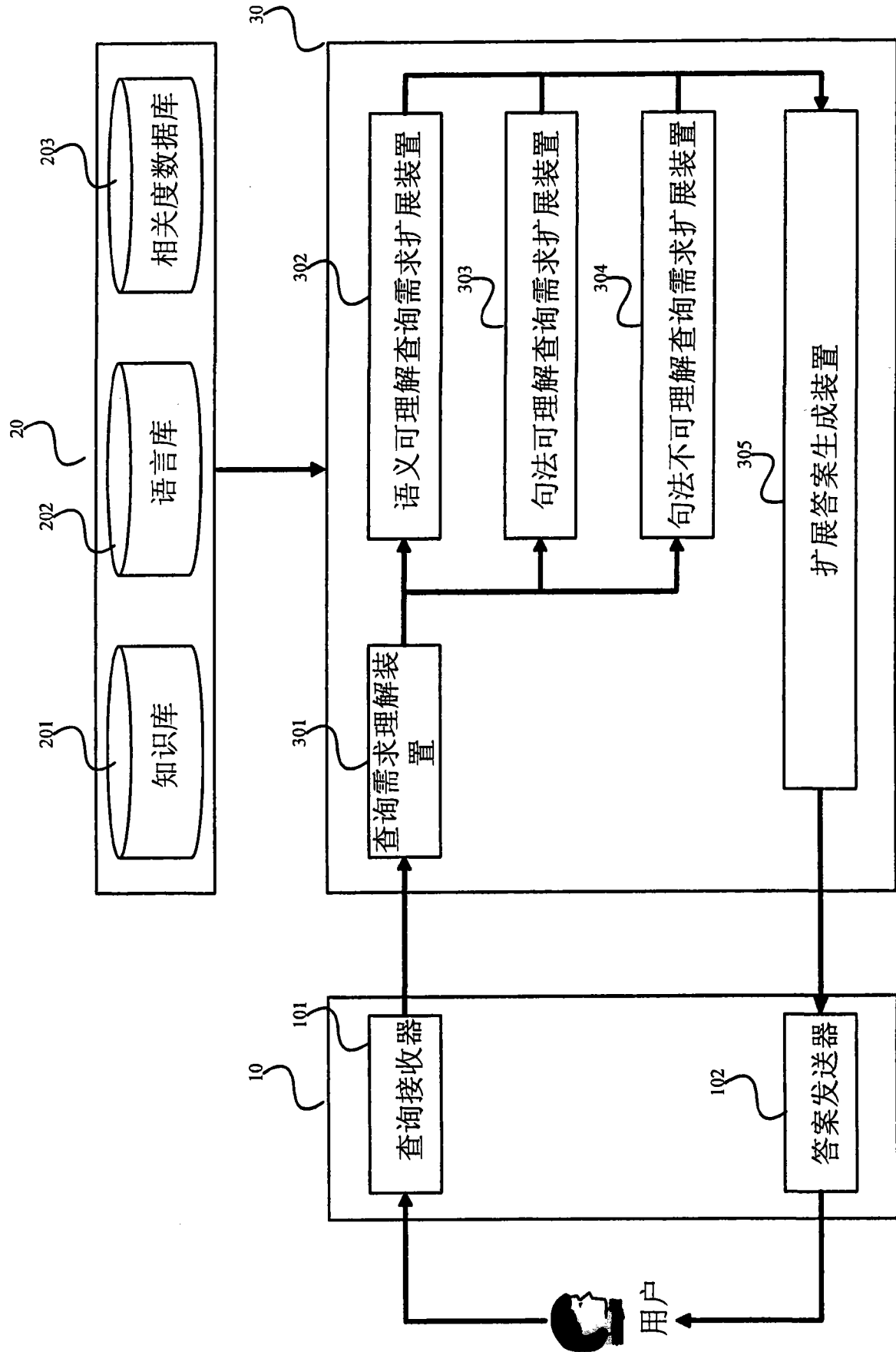


图 2

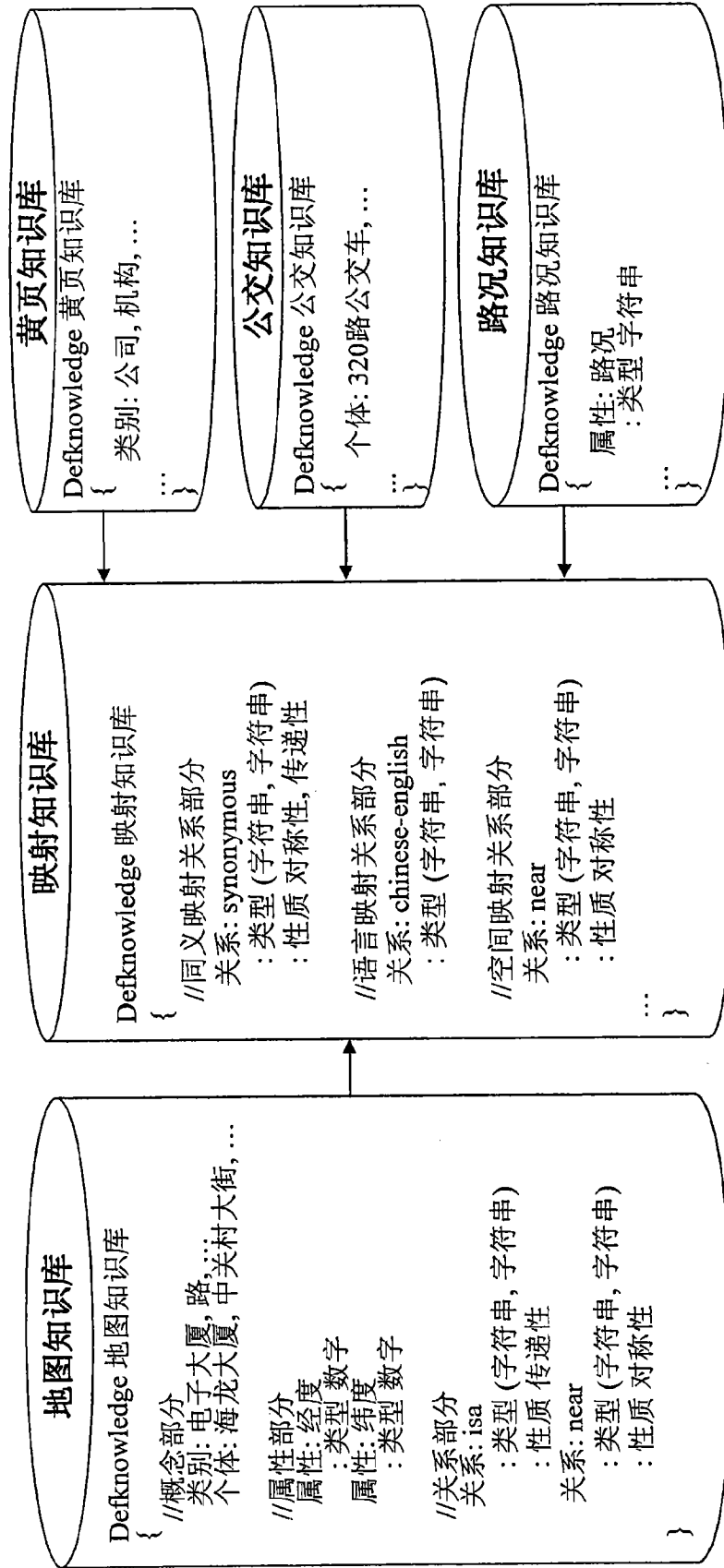


图 3(a)

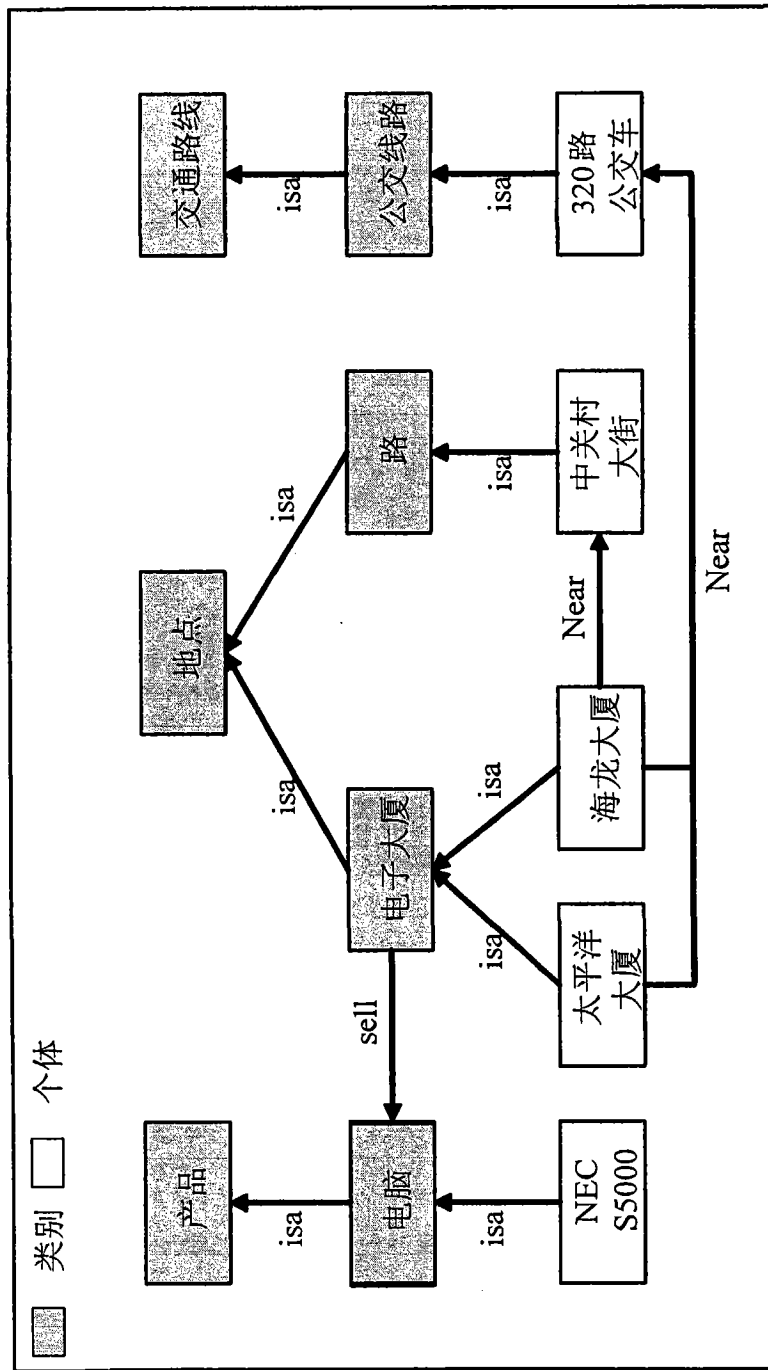


图 3 (b)

公用查询语言

```

DefSyntax 属性值查询
{
  文法: <?C(?cate)>[的]<?A>
}

```

路况领域查询语言

```

DefSyntax 路况: 属性值查询(?cate=路|桥, ?A=路况)
{
  文法: <?C(路|桥)><堵|拥堵|畅通>
  动作: 路况查询(?C)
}

```

公交领域查询语言

```

DefSyntax 路径: 属性值查询(?cate=地点, ?A=路径)
{
  文法: <如何|怎样|怎么><去|到|到达><?C(地点)>
  动作: 路径查询(?C)
}

```

黄页领域查询语言

```

DefSyntax 地址: 属性值查询(?cate=地点, ?A=地址)
{
  文法: <?C(地点)><在><哪里|哪儿|哪里|什么地方>
  动作: 地址查询(?C)
}

```

图 4

名称1	名称2	相关类型	相关度
海龙大厦	中关村大街	语义相关	1
海龙大厦	中关村大街	统计相关	0.7
太平洋	太平洋大厦	字符串相似	0.6
地址	路径	服务相关	0.9
...

图 5a

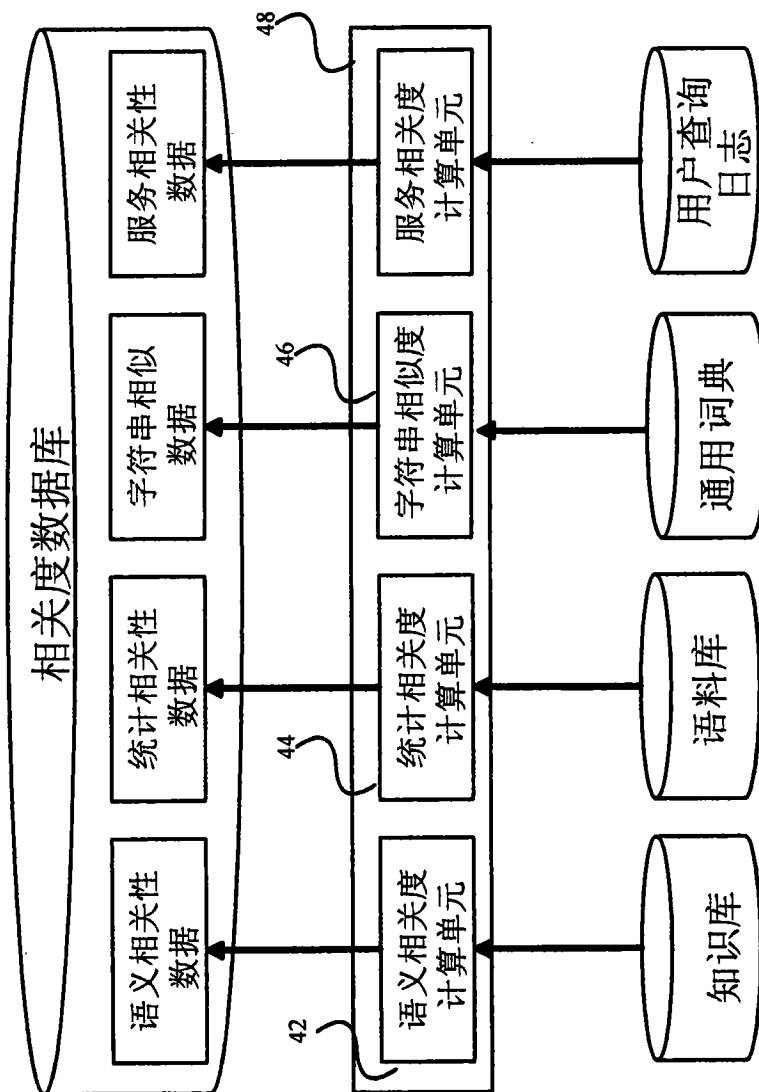


图 5b

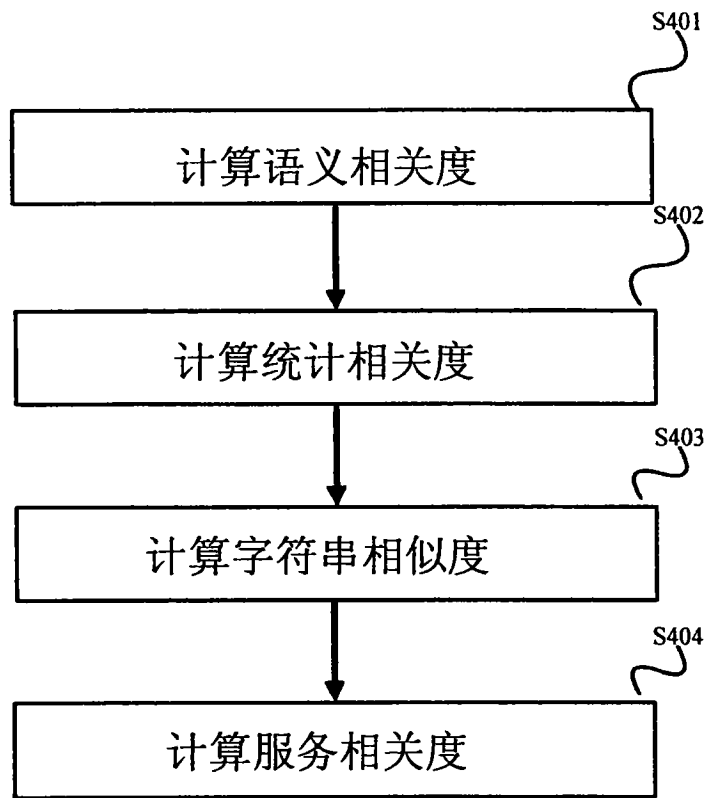


图 5c

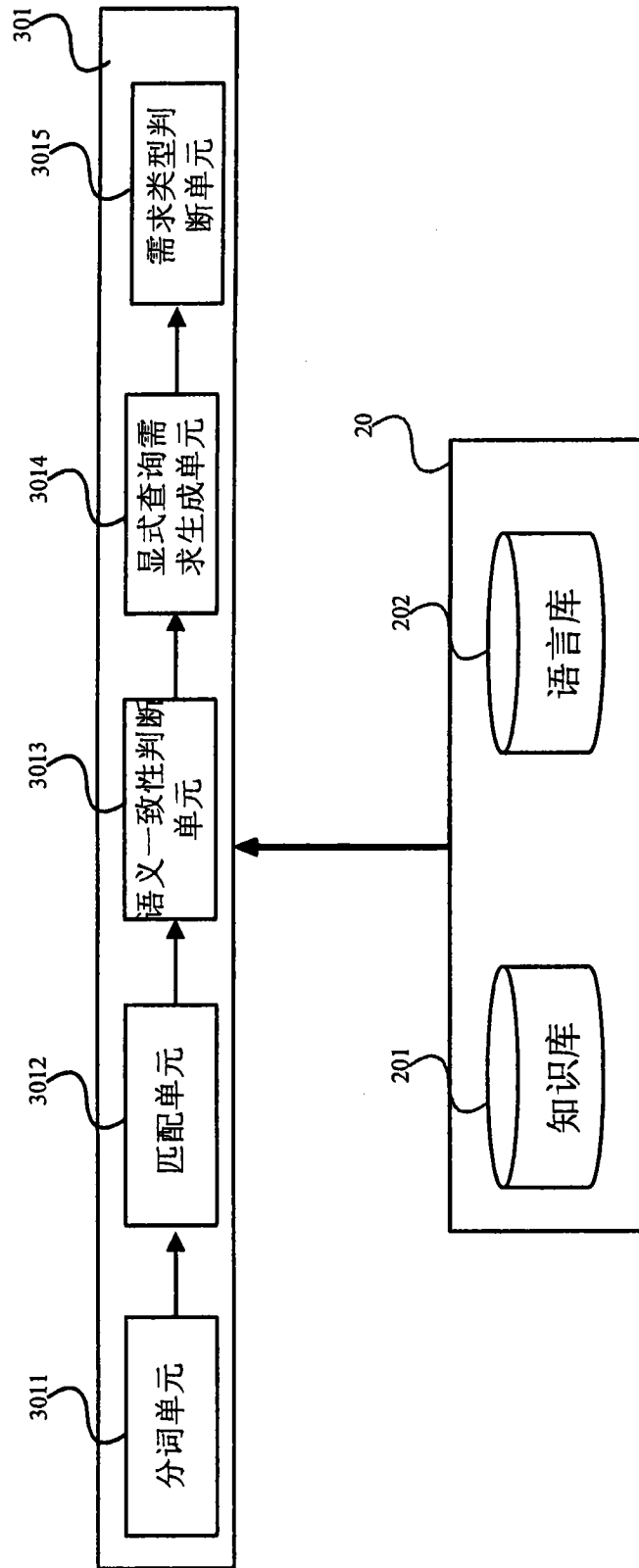


图 6a

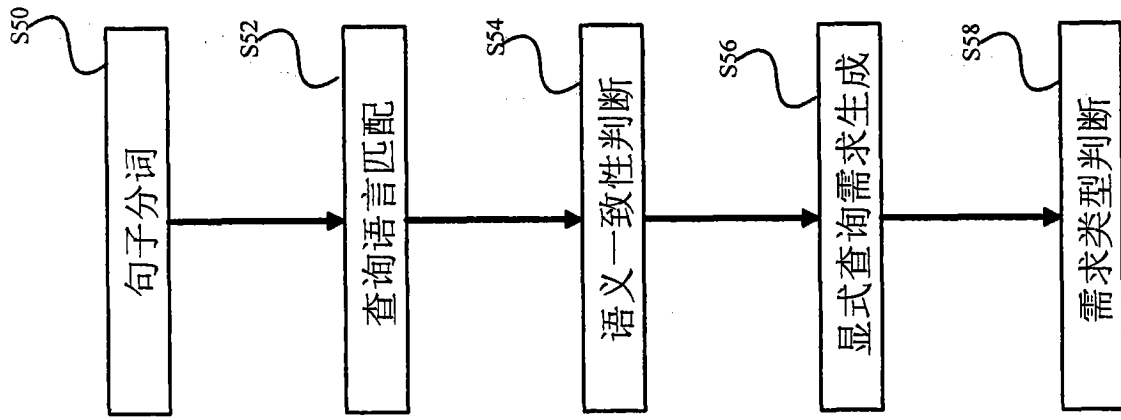


图 6b

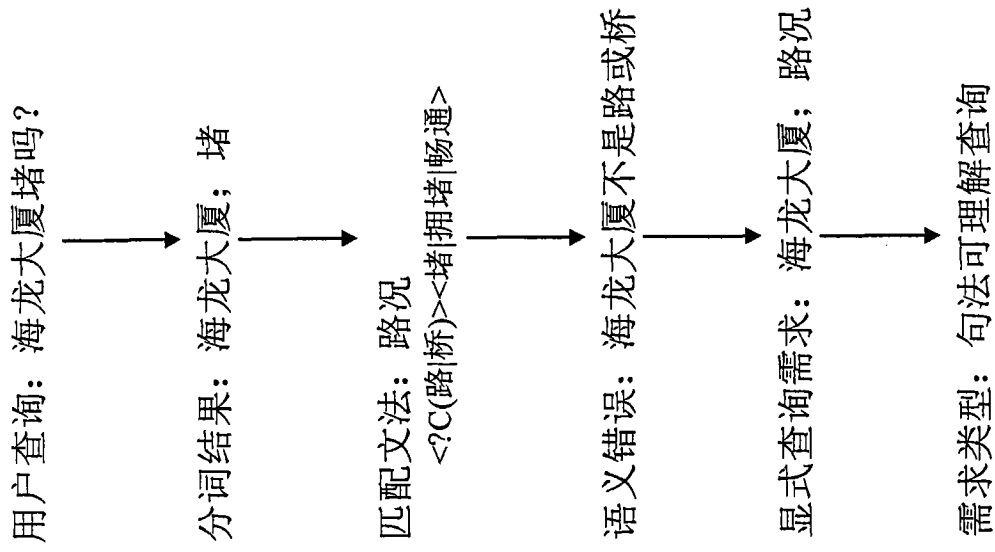


图 6c

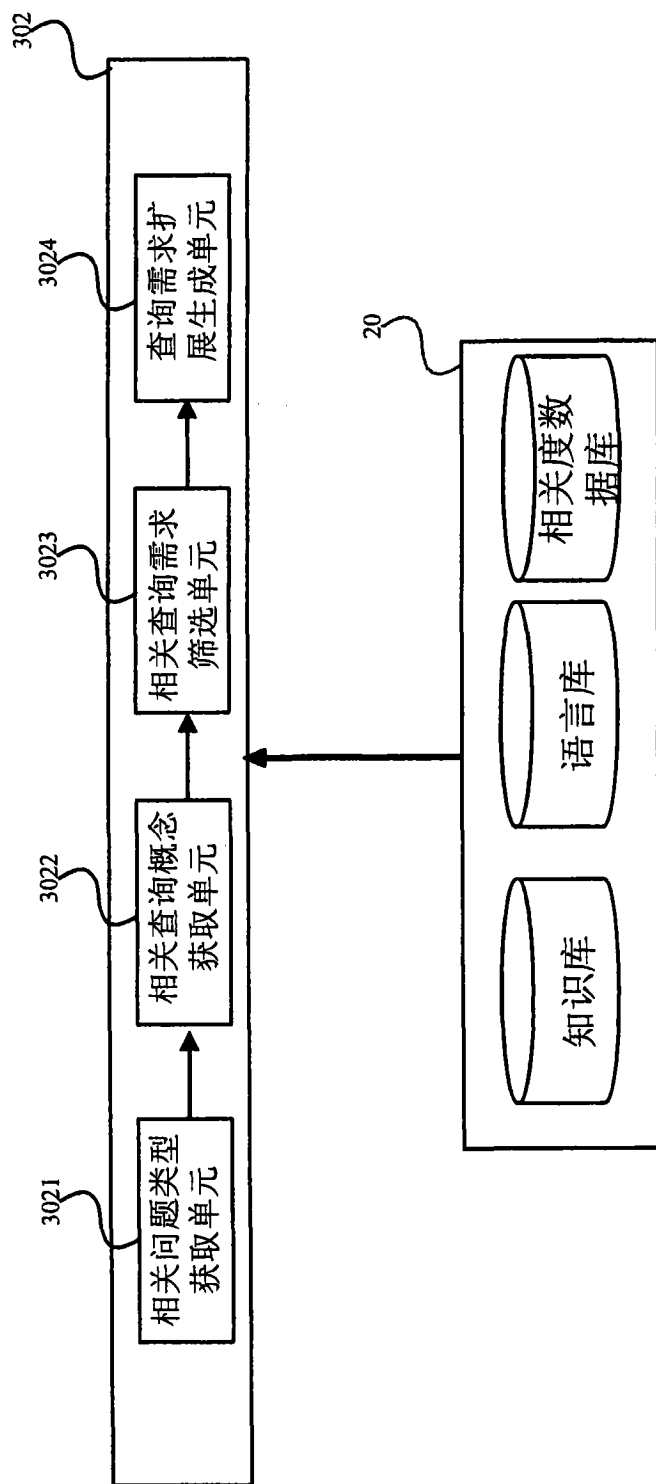


图 7a

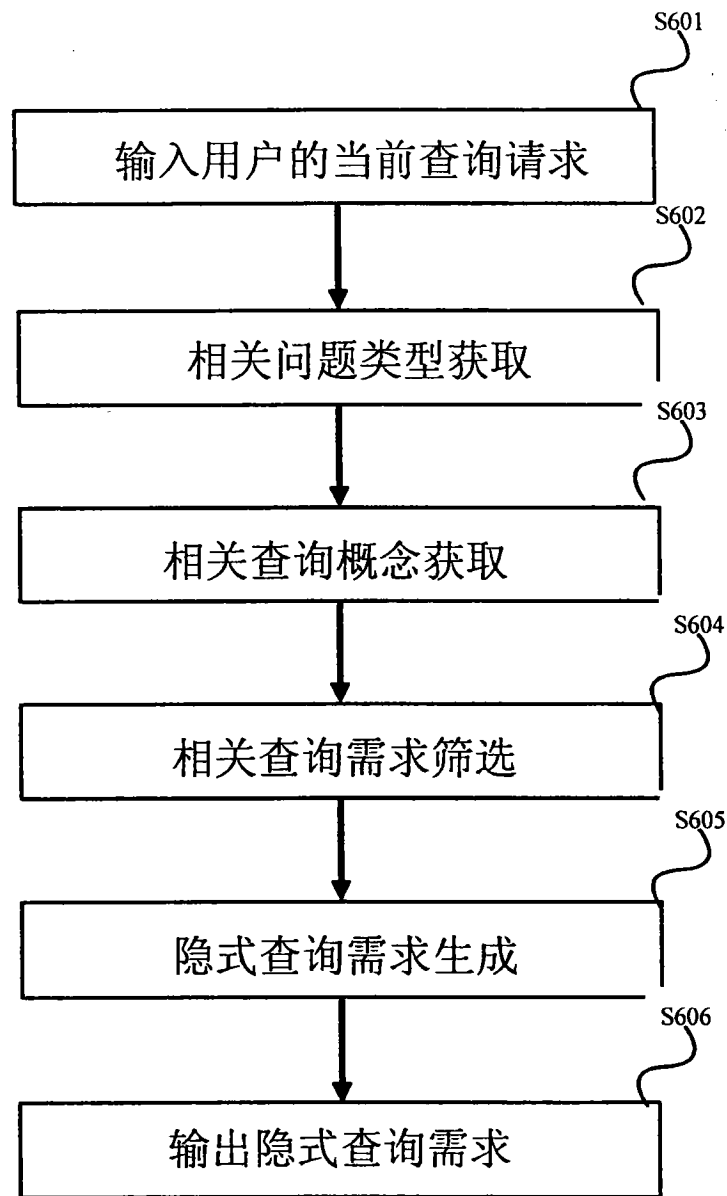


图 7b

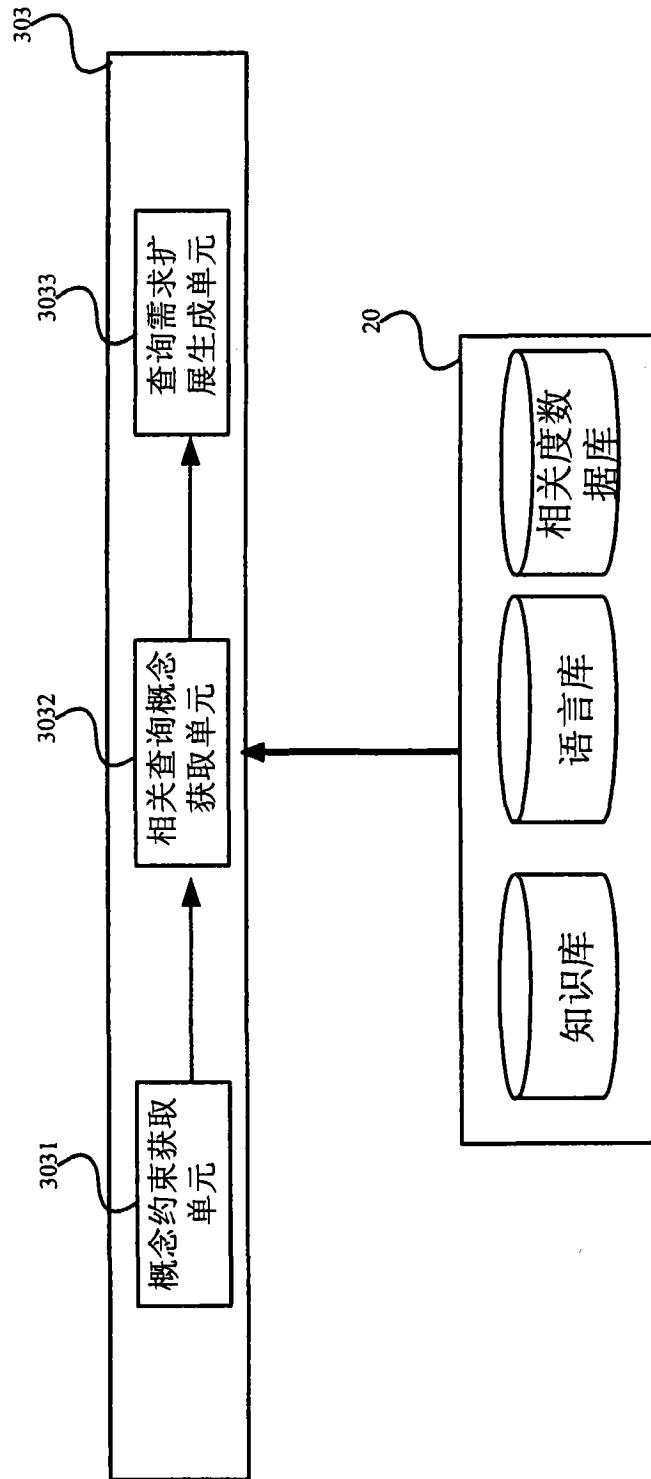


图 8a

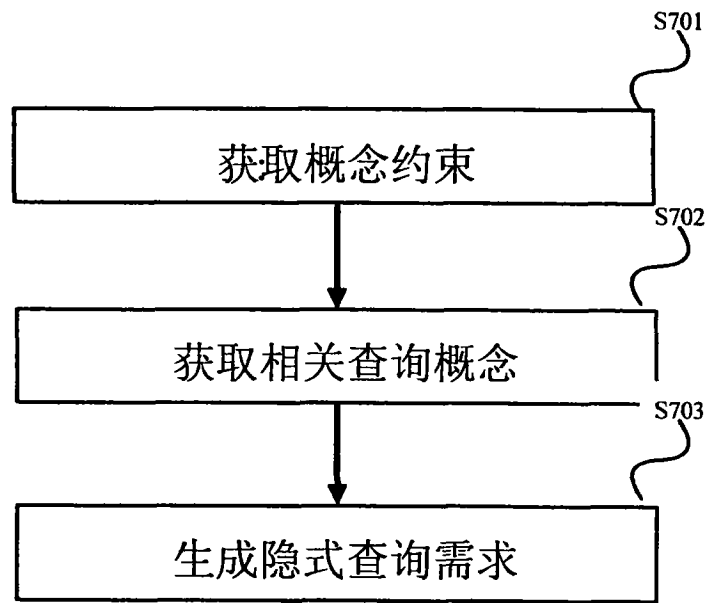


图 8b

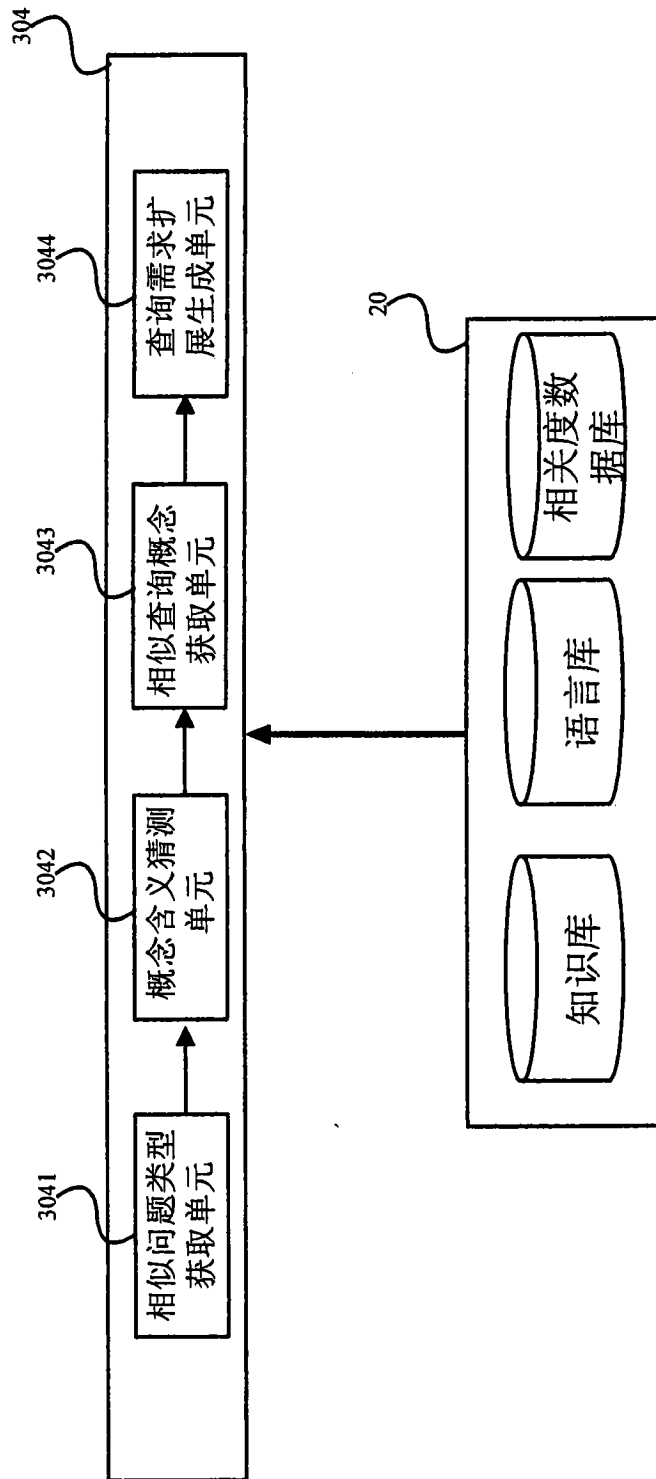


图 9a

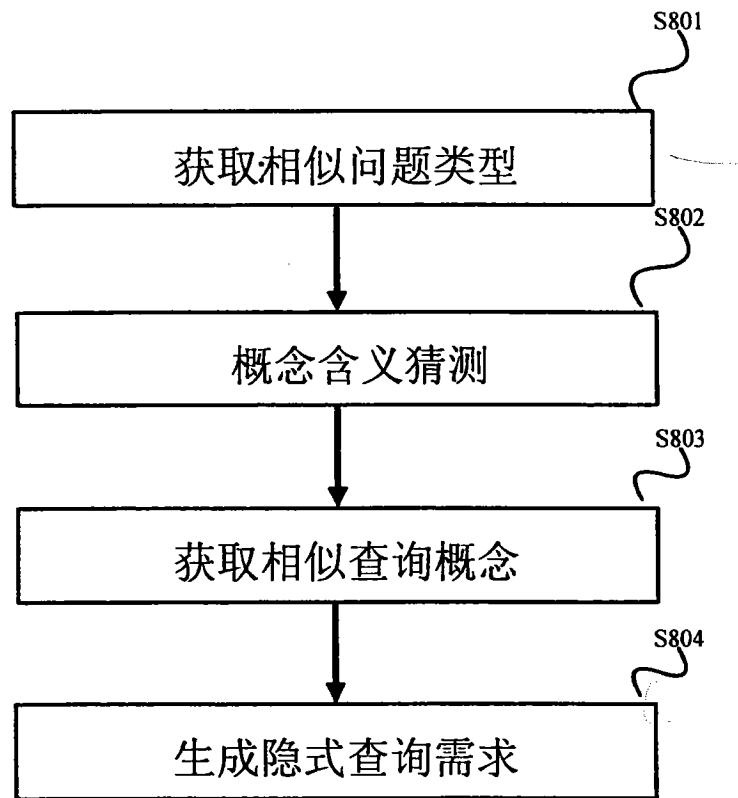


图 9b

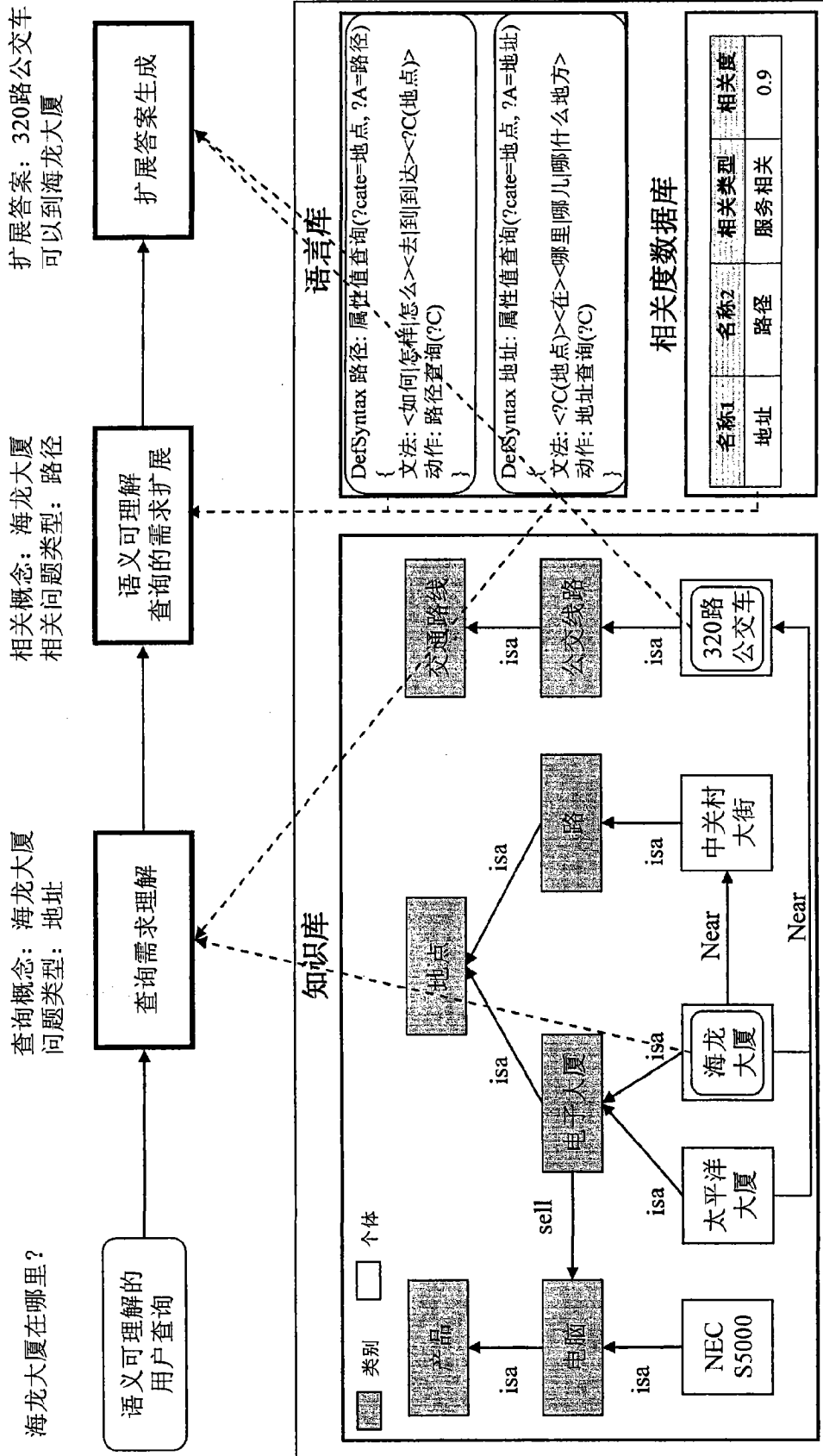


图 10a

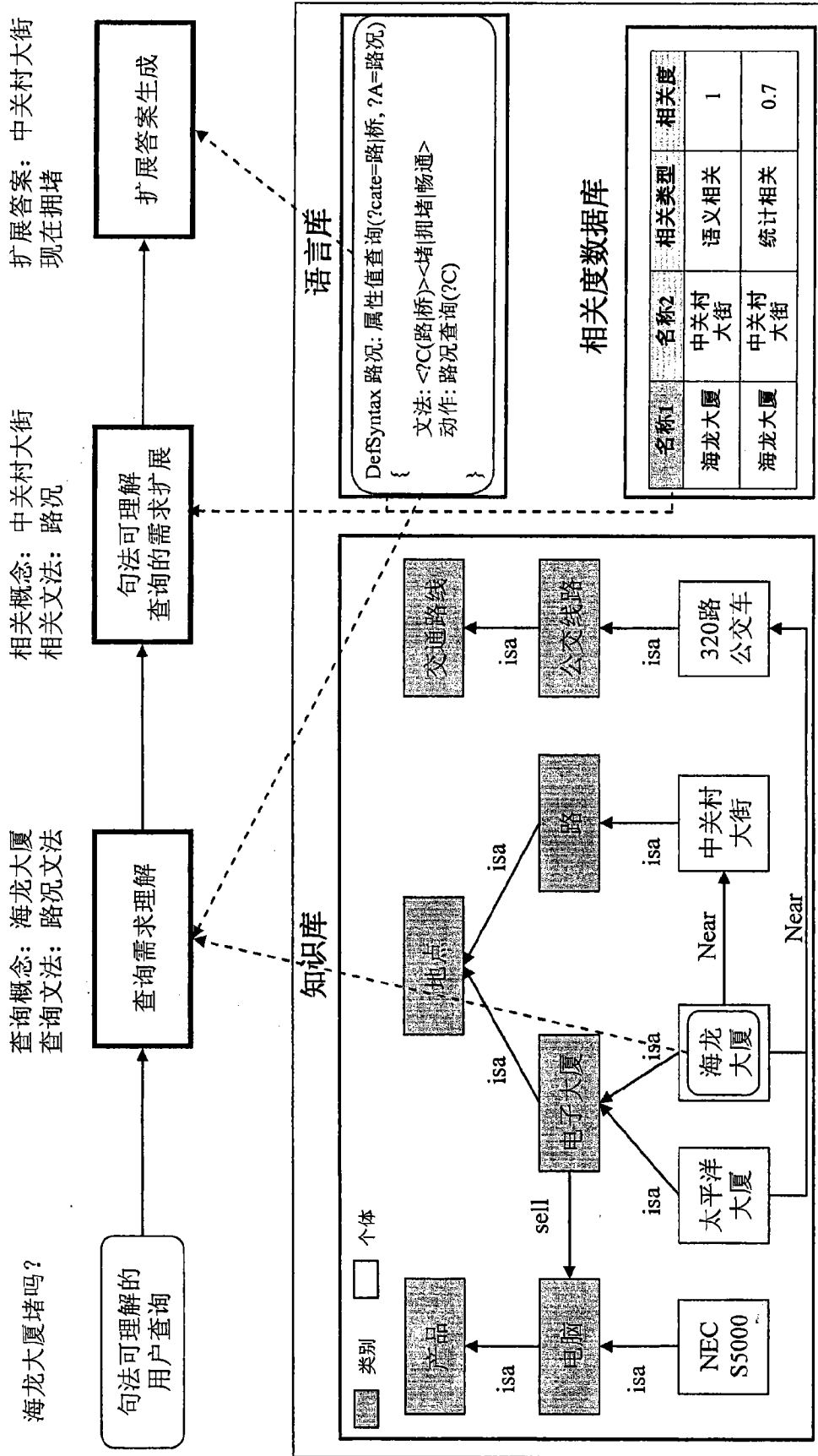


图 10b

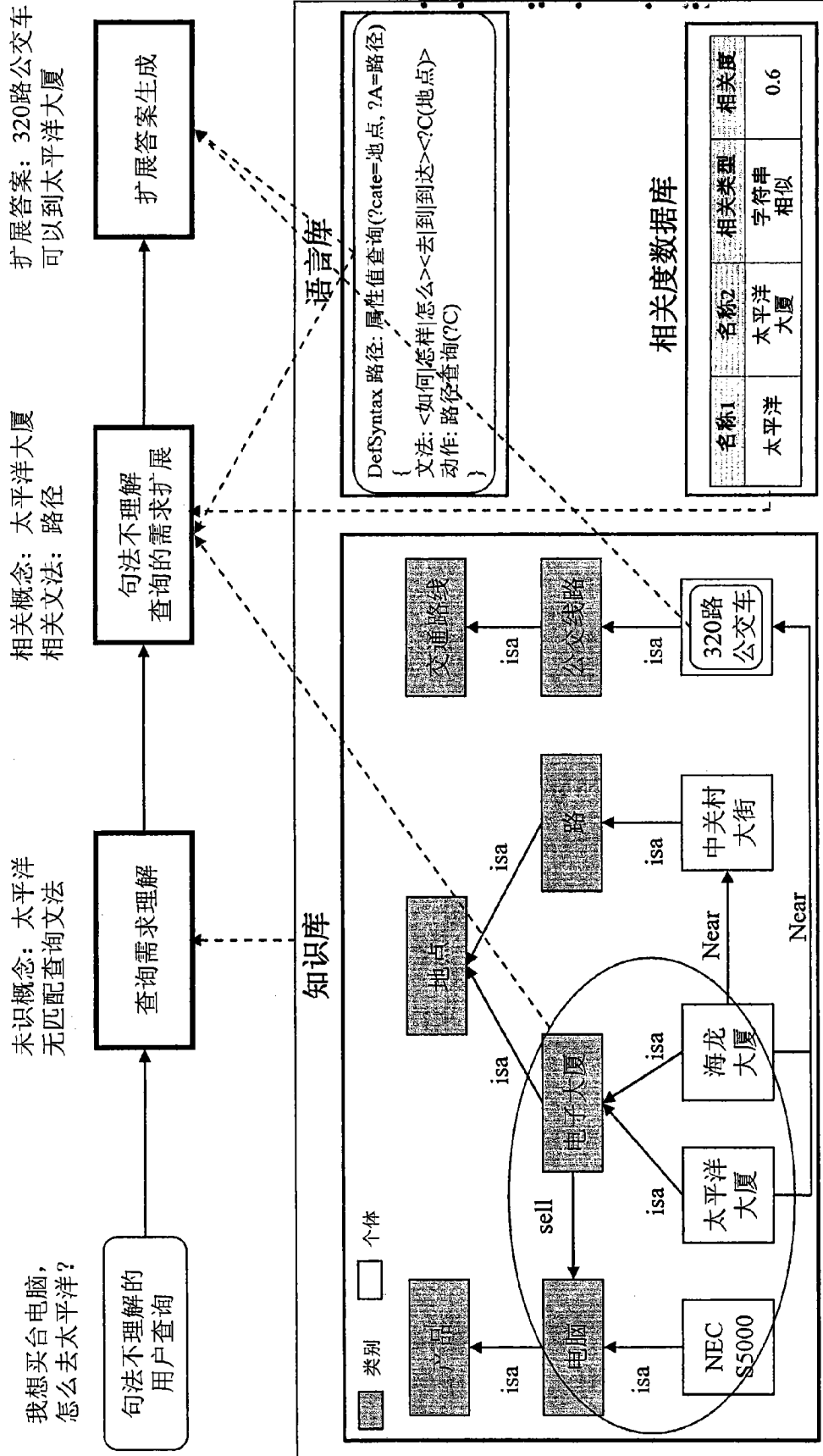


图 10c

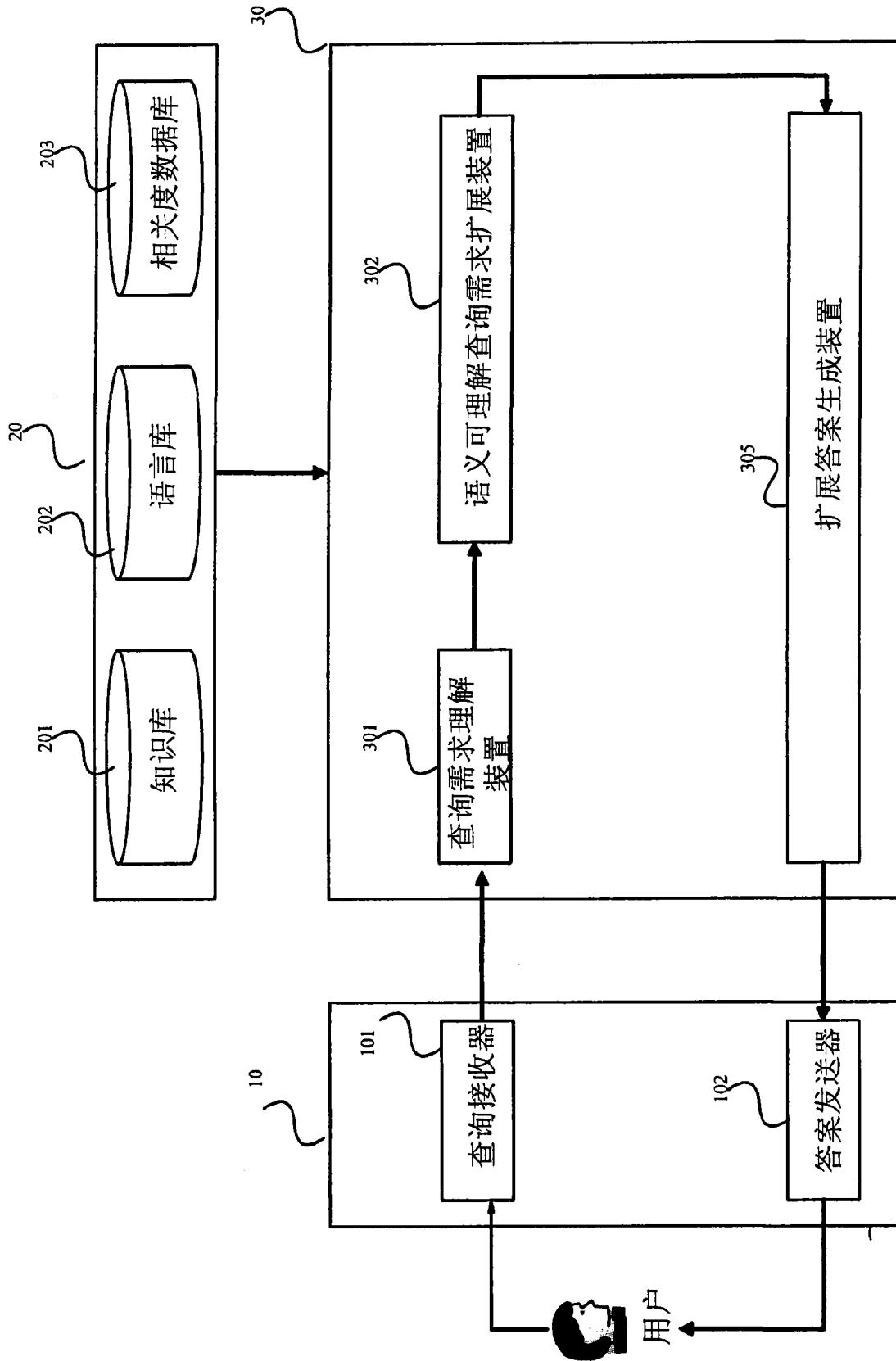


图 11