



(12) 发明专利申请

(10) 申请公布号 CN 112182150 A

(43) 申请公布日 2021.01.05

(21) 申请号 202011009559.2

(22) 申请日 2020.09.23

(71) 申请人 中国建设银行股份有限公司
地址 100033 北京市西城区金融大街25号

(72) 发明人 蒲柯锐 李昱 王全礼 王斌
孙昊

(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

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/335 (2019.01)

G06F 16/332 (2019.01)

G06F 16/901 (2019.01)

G06F 16/9035 (2019.01)

权利要求书4页 说明书14页 附图5页

(54) 发明名称

基于多元数据的聚合检索方法、装置、设备及存储介质

(57) 摘要

本发明公开了一种基于多元数据的聚合检索方法、装置、设备及存储介质。该方法包括：获取待检索的目标问题，并对所述目标问题进行意图识别处理，得到权重分词；采用多种检索引擎对所述权重分词进行检索，得到与各检索引擎对应的检索结果，所述检索引擎包括：与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎，以及与非结构化文档数据匹配的通用检索引擎；采用聚合引擎，根据各权重分词的权重，对与各检索引擎对应的检索结果进行置信度打分，并将得分大于预设阈值的检索结果顺序展示给用户。本发明实施例的技术方案，实现对多种类型数据进行聚合检索，扩大检索数据的支持范围，提高检索的准确度。



1. 一种基于多元数据的聚合检索方法,其特征在于,包括:

获取待检索的目标问题,并对所述目标问题进行意图识别处理,得到权重分词;

采用多种检索引擎对所述权重分词进行检索,得到与各检索引擎对应的检索结果,所述检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;

采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

2. 根据权利要求1所述的方法,其特征在于,获取待检索的目标问题,并对所述目标问题进行意图识别处理,得到权重分词,包括:

获取待检索的目标问题,并对所述目标问题进行归一化处理,得到初始分词,所述归一化处理包括:词过滤处理、词映射处理以及设置通用权重;

对各初始分词进行句法分析,根据各初始分词的句子结构筛选出有效分词;

从所述有效分词中提取实体以及实体属性,根据实体类别以及实体属性的词性为各有效分词设置图谱权重,得到权重分词。

3. 根据权利要求1所述的方法,其特征在于,采用多种检索引擎对所述权重分词进行检索,得到与各检索引擎对应的检索结果,包括:

调用所述问答引擎的接口,在问答对数据库中检索与所述权重分词匹配的问题,并将与所述问题匹配的答案作为检索结果;

调用所述图谱推理引擎的接口,根据与所述权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;

调用所述通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

4. 根据权利要求3所述的方法,其特征在于,调用所述图谱推理引擎的接口,根据与所述权重分词对应的图谱查询语句查询图谱数据库,得到检索结果,包括:

调用所述图谱推理引擎的接口,获取所述权重分词中的目标实体以及目标实体属性;

根据所述目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

5. 根据权利要求4所述的方法,其特征在于,根据所述目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据作为检索结果,包括:

如果所述目标实体没有所述目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体;

根据所述节点实体与所述目标实体属性生成图谱查询语句,并在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

6. 根据权利要求4所述的方法,其特征在于,在根据所述目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据之后,还包括:

如果所述权重分词中存在目标实体属性的参数值,则根据所述参数值,对所述至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推

理操作,得到与所述图谱推理引擎对应的检索结果。

7. 根据权利要求3所述的方法,其特征在于,调用所述通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果,包括:

调用所述通用检索引擎的接口,根据预设的关联词词库对所述权重分词进行关联词替换,并为替换后的权重分词设置通用权重;

根据所述权重分词,在预设的索引文件中进行倒排索引,得到索引数据;

根据排序模型,对所述索引数据进行二次排序,得到与所述通用检索引擎对应的检索结果。

8. 根据权利要求7所述的方法,其特征在于,还包括:

获取数据源的更新数据;

对所述更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;

从所述更新数据中提取实体以及实体属性,并在所述实体以及实体属性之间建立关联关系;

根据所述分类标签以及关联关系,将所述更新数据更新到所述预设的索引文件中。

9. 根据权利要求8所述的方法,其特征在于,获取数据源的更新数据,包括:

根据数据库归档日志,从与数据源对应的数据库中获取所述数据源的更新数据;或者

通过所述数据源提供的接口,获取所述数据源的更新数据;或者

根据所述数据源发送的批量文件,获取所述数据源的更新数据。

10. 根据权利要求7所述的方法,其特征在于,所述排序模型包括梯度提升树。

11. 根据权利要求1所述的方法,其特征在于,在采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,还包括:

采用半监督学习的方式,从非结构化文档数据中提取目标信息,所述目标信息包括实体、实体属性以及分类标签;

根据所述目标信息对结构化图谱数据进行归并处理,并对所述目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;

根据所述更新后的结构化图谱数据,为所述检索结果提供多维度关联信息。

12. 一种基于多元数据的聚合检索装置,其特征在于,包括:

获取模块,用于获取待检索的目标问题,并对所述目标问题进行意图识别处理,得到权重分词;

检索模块,用于采用多种检索引擎对所述权重分词进行检索,得到与各检索引擎对应的检索结果,所述检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;

打分模块,用于采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

13. 根据权利要求12所述的装置,其特征在于,获取模块,用于:

获取待检索的目标问题,并对所述目标问题进行归一化处理,得到初始分词,所述归一化处理包括:词过滤处理、词映射处理以及设置通用权重;

对各初始分词进行句法分析,根据各初始分词的句子结构筛选出有效分词;

从所述有效分词中提取实体以及实体属性,根据实体类别以及实体属性的词性为各有效分词设置图谱权重,得到权重分词。

14. 根据权利要求12所述的装置,其特征在于,检索模块,包括:

问答检索单元,用于调用所述问答引擎的接口,在问答对数据库中检索与所述权重分词匹配的问题,并将与所述问题匹配的答案作为检索结果;

图谱检索单元,用于调用所述图谱推理引擎的接口,根据与所述权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;

通用检索单元,用于调用所述通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

15. 根据权利要求14所述的装置,其特征在于,图谱检索单元,包括:

获取子单元,用于调用所述图谱推理引擎的接口,获取所述权重分词中的目标实体以及目标实体属性;

查询子单元,用于根据所述目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

16. 根据权利要求15所述的装置,其特征在于,查询子单元,用于:

如果所述目标实体没有所述目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体;

根据所述节点实体与所述目标实体属性生成图谱查询语句,并在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

17. 根据权利要求15所述的装置,其特征在于,查询子单元,还用于:

在根据所述目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与所述图谱查询语句匹配的至少一项结构化图谱数据之后,

如果所述权重分词中存在目标实体属性的参数值,则根据所述参数值,对所述至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,得到与所述图谱推理引擎对应的检索结果。

18. 根据权利要求14所述的装置,其特征在于,通用检索单元,包括:

替换子单元,用于调用所述通用检索引擎的接口,根据预设的关联词词库对所述权重分词进行关联词替换,并为替换后的权重分词设置通用权重;

索引子单元,用于根据所述权重分词,在预设的索引文件中进行倒排索引,得到索引数据;

排序子单元,用于根据排序模型,对所述索引数据进行二次排序,得到与所述通用检索引擎对应的检索结果。

19. 根据权利要求18所述的装置,其特征在于,还包括:更新模块,用于:

获取数据源的更新数据;

对所述更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;

从所述更新数据中提取实体以及实体属性,并在所述实体以及实体属性之间建立关联关系;

根据所述分类标签以及关联关系,将所述更新数据更新到所述预设的索引文件中。

20. 根据权利要求12所述的装置,其特征在于,还包括:

多维度关联模块,用于:在采用聚合引擎,根据各权重分词的权重,对各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,

采用半监督学习的方式,从非结构化文档数据中提取目标信息,所述目标信息包括实体、实体属性以及分类标签;

根据所述目标信息对结构化图谱数据进行归并处理,并对所述目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;

根据所述更新后的结构化图谱数据,为所述检索结果提供多维度关联信息。

21. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-11中任一所述的基于多元数据的聚合检索方法。

22. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1-11中任一所述的基于多元数据的聚合检索方法。

基于多元数据的聚合检索方法、装置、设备及存储介质

技术领域

[0001] 本发明实施例涉及计算机技术领域,尤其涉及一种基于多元数据的聚合检索方法、装置、设备及存储介质。

背景技术

[0002] 随着各个领域信息量的不断增加,用户对信息检索的需求也日益增加。目前,大多数检索引擎都是基于传统的检索技术构建的,存在信息过载,返回的检索结果过多,有很多与检索意图无关的检索结果的问题,对用户实际的搜索意图缺乏准确的判断,检索的准确度不高。同时,检索结果形式单一,无法把垂直领域中多类多结构的相关数据都检索出来,导致用户二次搜索的频率较高。

发明内容

[0003] 本发明提供一种基于多元数据的聚合检索方法、装置、设备及存储介质,以实现多种类型数据进行聚合检索,扩大检索数据的支持范围,提高检索的准确度。

[0004] 第一方面,本发明实施例提供了一种基于多元数据的聚合检索方法,包括:

[0005] 获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;

[0006] 采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;

[0007] 采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0008] 可选的,获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词,包括:

[0009] 获取待检索的目标问题,并对目标问题进行归一化处理,得到初始分词,归一化处理包括:词过滤处理、词映射处理以及设置通用权重;

[0010] 对各初始分词进行句法分析,根据各初始分词的句子结构筛选出有效分词;

[0011] 从有效分词中提取实体以及实体属性,根据实体类别以及实体属性的词性为各有效分词设置图谱权重,得到权重分词。

[0012] 可选的,采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,包括:

[0013] 调用问答引擎的接口,在问答对数据库中检索与权重分词匹配的问题,并将与问题匹配的答案作为检索结果;

[0014] 调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;

[0015] 调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

- [0016] 可选的,调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果,包括:
- [0017] 调用图谱推理引擎的接口,获取权重分词中的目标实体以及目标实体属性;
- [0018] 根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。
- [0019] 可选的,根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果,包括:
- [0020] 如果目标实体没有目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体;
- [0021] 根据节点实体与目标实体属性生成图谱查询语句,并在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。
- [0022] 可选的,在根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据之后,还包括:
- [0023] 如果权重分词中存在目标实体属性的参数值,则根据参数值,对至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,得到与图谱推理引擎对应的检索结果。
- [0024] 可选的,调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果,包括:
- [0025] 调用通用检索引擎的接口,根据预设的关联词词库对权重分词进行关联词替换,并为替换后的权重分词设置通用权重;
- [0026] 根据权重分词,在预设的索引文件中进行倒排索引,得到索引数据;
- [0027] 根据排序模型,对索引数据进行二次排序,得到与通用检索引擎对应的检索结果。
- [0028] 可选的,还包括:
- [0029] 获取数据源的更新数据;
- [0030] 对更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;
- [0031] 从更新数据中提取实体以及实体属性,并在实体以及实体属性之间建立关联关系;
- [0032] 根据分类标签以及关联关系,将更新数据更新到预设的索引文件中。
- [0033] 可选的,获取数据源的更新数据,包括:
- [0034] 根据数据库归档日志,从与数据源对应的数据库中获取数据源的更新数据;或者
- [0035] 通过数据源提供的接口,获取数据源的更新数据;或者
- [0036] 根据数据源发送的批量文件,获取数据源的更新数据。
- [0037] 可选的,排序模型包括梯度提升树。
- [0038] 可选的,在采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,还包括:
- [0039] 采用半监督学习的方式,从非结构化文档数据中提取目标信息,目标信息包括实体、实体属性以及分类标签;
- [0040] 根据目标信息对结构化图谱数据进行归并处理,并对目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;

- [0041] 根据更新后的结构化图谱数据,为检索结果提供多维度关联信息。
- [0042] 第二方面,本发明实施例还提供了一种基于多元数据的聚合检索装置,包括:
- [0043] 获取模块,用于获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;
- [0044] 检索模块,用于采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;
- [0045] 打分模块,用于采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。
- [0046] 可选的,获取模块,用于:
- [0047] 获取待检索的目标问题,并对目标问题进行归一化处理,得到初始分词,归一化处理包括:词过滤处理、词映射处理以及设置通用权重;
- [0048] 对各初始分词进行句法分析,根据各初始分词的句子结构筛选出有效分词;
- [0049] 从有效分词中提取实体以及实体属性,根据实体类别以及实体属性的词性为各有效分词设置图谱权重,得到权重分词。
- [0050] 可选的,检索模块,包括:
- [0051] 问答检索单元,用于调用问答引擎的接口,在问答对数据库中检索与权重分词匹配的问题,并将与问题匹配的答案作为检索结果;
- [0052] 图谱检索单元,用于调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;
- [0053] 通用检索单元,用于调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。
- [0054] 可选的,图谱检索单元,包括:
- [0055] 获取子单元,用于调用图谱推理引擎的接口,获取权重分词中的目标实体以及目标实体属性;
- [0056] 查询子单元,用于根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。
- [0057] 可选的,查询子单元,用于:
- [0058] 如果目标实体没有目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体;
- [0059] 根据节点实体与目标实体属性生成图谱查询语句,并在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。
- [0060] 可选的,查询子单元,还用于:
- [0061] 在根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据之后,
- [0062] 如果权重分词中存在目标实体属性的参数值,则根据参数值,对至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,得到与图谱推理引擎对应的检索结果。
- [0063] 可选的,通用检索单元,包括:

- [0064] 替换子单元,用于调用通用检索引擎的接口,根据预设的关联词词库对权重分词进行关联词替换,并为替换后的权重分词设置通用权重;
- [0065] 索引子单元,用于根据权重分词,在预设的索引文件中进行倒排索引,得到索引数据;
- [0066] 排序子单元,用于根据排序模型,对索引数据进行二次排序,得到与通用检索引擎对应的检索结果。
- [0067] 可选的,更新模块,用于:
- [0068] 获取数据源的更新数据;
- [0069] 对更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;
- [0070] 从更新数据中提取实体以及实体属性,并在实体以及实体属性之间建立关联关系;
- [0071] 根据分类标签以及关联关系,将更新数据更新到预设的索引文件中。
- [0072] 可选的,还包括:多维度关联模块,用于:在采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,
- [0073] 采用半监督学习的方式,从非结构化文档数据中提取目标信息,目标信息包括实体、实体属性以及分类标签;
- [0074] 根据目标信息对结构化图谱数据进行归并处理,并对目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;
- [0075] 根据更新后的结构化图谱数据,为检索结果提供多维度关联信息。
- [0076] 第三方面,本发明实施例还提供了一种电子设备,包括:
- [0077] 一个或多个处理器;
- [0078] 存储装置,用于存储一个或多个程序,
- [0079] 当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现本发明任意实施例提供的基于多元数据的聚合检索方法。
- [0080] 第四方面,本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现本发明任意实施例提供的基于多元数据的聚合检索方法。
- [0081] 本发明实施例的技术方案,通过获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户,解决了现有技术中存在的检索准确度不高以及二次搜索频繁的问题,实现对多种类型数据进行聚合检索,扩大检索数据的支持范围,提高检索的准确度。

附图说明

- [0082] 图1a是本发明实施例一中的一种基于多元数据的聚合检索方法的流程图;
- [0083] 图1b是本发明实施例一中的一种意图识别的实现流程图;

- [0084] 图1c是本发明实施例一中的一种聚合检索方法的流程图；
- [0085] 图1d是本发明实施例一中的一种对检索结果进行信息关联的实现流程图；
- [0086] 图2a是本发明实施例二中的一种基于多元数据的聚合检索方法的流程图；
- [0087] 图2b是本发明实施例二中的一种通用检索引擎的检索流程图；
- [0088] 图3是本发明实施例三中的一种基于多元数据的聚合检索装置的结构示意图；
- [0089] 图4是本发明实施例四中的一种电子设备的结构示意图。

具体实施方式

[0090] 下面结合附图和实施例对本发明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，而非对本发明的限定。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相关的部分而非全部结构。

[0091] 实施例一

[0092] 图1a是本发明实施例一中的一种基于多元数据的聚合检索方法的流程图，本实施例可适用于基于多种类型的数据进行聚合检索的情况，该方法可以由基于多元数据的聚合检索装置来执行，该装置可以由硬件和/或软件来实现，并一般可以集成在提供数据检索服务的电子设备中。如图1a所示，该方法包括：

[0093] 步骤110、获取待检索的目标问题，并对目标问题进行意图识别处理，得到权重分词。

[0094] 本实施例中，目标问题可以是用户输入的等待检索的问题，为了能准确获取到用户想要检索的内容，可以先对目标问题进行意图识别处理，删除目标问题中与检索意图无关的内容，筛选出可以体现用户检索意图的内容，以便于后续进行检索。

[0095] 可选的，获取待检索的目标问题，并对目标问题进行意图识别处理，得到权重分词，可以包括：获取待检索的目标问题，并对目标问题进行归一化处理，得到初始分词，归一化处理包括：词过滤处理、词映射处理以及设置通用权重；对各初始分词进行句法分析，根据各初始分词的句子结构筛选出有效分词；从有效分词中提取实体以及实体属性，根据实体类别以及实体属性的词性为各有效分词设置图谱权重，得到权重分词。

[0096] 本实施例中，如图1b所示，对目标问题进行意图识别处理时，可以先对目标问题进行归一化处理，包括：过滤目标问题中的停用词、客气词、虚拟词等无实际意义的分词，例如，过滤目标问题“信用卡的申办条件是什么呀？”中的“的”、“是什么”、“呀”、“？”等分词；然后，对目标问题中的剩余分词进行中英文映射、缩略词映射、同音词纠错等词映射处理，例如，将英文分词转换成中文分词，将缩略语转换成中文全拼等；然后，根据预先设置的垂直领域业务词库，为各分词设置针对通用检索引擎的通用权重，得到初始分词。其中，为在业务词库中的分词设置较大的权重，为不在业务词库中的分词设置较小的权重，通用权重可以影响与通用检索引擎对应的检索结果的最终筛选和排序。

[0097] 然后，对各初始分词进行句法分析，即，分析出各初始分词对应的句子结构，例如，主语、谓语、宾语、状语等，并获取与各句子结构对应的词语属性，例如，名词、动词、形容词、代词、数词等，进而可以根据句子结构以及与句子结构对应的词语属性进一步筛选出初始分词中的有效分词；然后，可以从有效分词中提取实体以及实体属性，例如，实体为红十字会会员龙卡，实体属性为卡样，则实体类别为信用卡，实体属性的词性为名词，则可以根据实

体类别以及实体属性的词性为有效分词“卡样”设置较小的图谱权重,得到权重分词。

[0098] 其中,也可以为实体属性设置等级,例如,对于信用卡实体,每种信用卡的申请和挂失都是比较相似的,因此,可以为申请和挂失这种不同信用卡之间比较相似的实体属性设置较低的等级,每种信用卡的年费是不同的,因此,可以为年费这种不同信用卡之间差距较大的实体属性设置较高的等级。进而,可以根据实体类别以及实体属性的等级,为等级较高的有效分词设置较大的图谱权重,为等级较低的有效分词设置较小的图谱权重,得到权重分词。

[0099] 本实施例中,垂直领域业务词库中包括针对目标行业领域预先设置的业务分词,以及针对与目标行业领域相关的垂直领域设置的业务分词,通过对在业务词库中的分词设置较大的通用权重,可以使得与目标行业领域专业相关的检索结果可以被展现给用户,提高检索的专业度。通过对各有效分词设置图谱权重,使得图谱推理引擎经过推理检索得到的有效检索结果可以被展现给用户,提高检索准确度。

[0100] 步骤120、采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果。

[0101] 其中,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎。

[0102] 本实施例中,如图1c所示,为了可以扩大检索数据的支持范围,支持同时对多种类型的数据进行检索,将垂直领域中多类多结构的相关数据都检索出来,设置了与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎,进而可以采用问答引擎、图谱推理引擎以及通用检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果。

[0103] 可选的,采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,可以包括:调用问答引擎的接口,在问答对数据库中检索与权重分词匹配的问题,并将与问题匹配的答案作为检索结果;调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

[0104] 本实施例中,如图1c所示,针对用户输入的目标问题,可以同时采用问答引擎、图谱推理引擎以及通用检索引擎对目标问题进行检索,以便于从多种数据类型的检索结果中选择与目标问题精准匹配的问题答案。可以调用问答引擎的接口,从问答对数据库中查询与权重分词匹配的问题,并将与问题匹配的答案作为问答对类型的检索结果;调用图谱推理引擎的接口,根据权重分词生成对应的图谱查询语句,并使用图谱查询语句在图谱数据库中查询到结构化图谱数据类型的检索结果;调用通用检索引擎的接口,根据权重分词在预设的索引文件中进行索引,得到非结构化文档数据类型的检索结果。

[0105] 步骤130、采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0106] 本实施例中,如图1c所示,在得到多种数据类型的检索结果之后,通过调用聚合引擎的接口,根据各权重分词对应的权重,对与各检索引擎对应的检索结果进行置信度打分,其中,权重较大的权重分词对应的检索结果的得分较高,将得分大于预设阈值的检索结果筛选出来作为与目标问题匹配的检索结果,并将筛选出的所有检索结果按照得分由高到低

的顺序展示给用户。

[0107] 可选的,在采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,还可以包括:采用半监督学习的方式,从非结构化文档数据中提取目标信息,目标信息包括实体、实体属性以及分类标签;根据目标信息对结构化图谱数据进行归并处理,并对目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;根据更新后的结构化图谱数据,为检索结果提供多维度关联信息。

[0108] 本实施例中,如图1d所示,在向用户展示检索结果之后,为了能够对检索结果关联多维度信息,减少用户基于检索结果进行二次检索的频率,可以采用半监督学习的方式,从非结构化文档数据中提取实体、实体属性以及对应的分类标签,然后将提取的实体、实体属性以及对应的分类标签归并融合到结构化图谱数据中,即与图谱数据进行相似信息融合,以完善补充图谱数据。再用完善后的结构化图谱数据为检索结果提供多维度关联信息,例如,对于信用卡信息文档,对文档中涉及的专有名词提供名词解释关联;对于基金类资讯文档,对文档中涉及的基金产品进行信息关联;对于理财产品说明文档,基于文档内容进行相似文档关联等。

[0109] 本发明实施例的技术方案,通过获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户,解决了现有技术中存在的检索准确度不高以及二次搜索频繁的问题,实现对多种类型数据进行聚合检索,扩大检索数据的支持范围,提高检索的准确度。

[0110] 实施例二

[0111] 图2a是本发明实施例二中的一种基于多元数据的聚合检索方法的流程图,本实施例在上述实施例的基础上进一步细化,提供了采用图谱推理引擎进行检索的具体步骤,以及采用通用检索引擎进行检索的具体步骤。下面结合图2a对本申请实施例二中提供的一种基于多元数据的聚合检索方法进行说明,包括以下步骤:

[0112] 步骤210、获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词。

[0113] 本实施例中,用户可以手动输入目标问题,也可以通过麦克风语音输入目标问题,或者可以从提供的问题选择列表中选择目标问题。

[0114] 步骤220、调用问答引擎的接口,在问答对数据库中检索与权重分词匹配的问题,并将与问题匹配的答案作为检索结果。

[0115] 本实施例中,对数据库中存储有若干对问题和答案,通过调用问答引擎的接口,可以根据权重分词,从问答对数据库中查询目标问题,如果查询到,则将与目标问题匹配的答案作为检索结果,如果查询不到目标问题,则将与目标问题需相近的问题的答案作为检索结果。

[0116] 步骤230、调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果。

[0117] 本实施例中,图谱数据库中既包括若干不同的结构化图谱数据,又包括各数据之间的关联关系,因此,根据图谱查询语句既可以检索到与目标问题匹配的答案,又可以检索到与目标问题相关联的其他内容。

[0118] 可选的,调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果,可以包括:调用图谱推理引擎的接口,获取权重分词中的目标实体以及目标实体属性;根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

[0119] 本实施例中,调用图谱推理引擎进行检索时,先提取权重分词中的目标实体以及目标实体属性,例如,权重分词“信用卡”为目标实体,“申办”和“条件”为目标实体属性,然后根据“信用卡”、“申办”和“条件”生成图谱查询语句,并在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据,需要说明的是,针对同一个目标问题的相同的权重分词,根据其中的目标实体属性是否足够具体,可能检索到与目标问题唯一匹配的答案,也可能检索到多个与目标问题匹配的答案。

[0120] 示例性的,对于权重分词“信用卡”、“申办”和“条件”,信用卡的申办条件已经足够具体并且申办条件是固定的,可以获取到与目标问题唯一匹配的答案;对于权重分词“信用卡”、“激活”,由于在手机银行、网银、网点、微信等不同渠道的信用卡激活方式不同,因此,会检索到与目标问题匹配的多个答案。

[0121] 可选的,根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果,可以包括:如果目标实体没有目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体,根据节点实体与目标实体属性生成图谱查询语句,并在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

[0122] 本实施例中,如果权重分词中的实体属性实际上为虚拟属性,例如,权重分词中的目标实体为“商旅航空系列”,目标实体属性为“卡样”,但是,对于实体“商旅航空系列”,其没有“卡样”这个实体属性,则可以定位到目标实体的下一层实体,如果其下一层实体也没有“卡样”属性,则继续定位到再下一层实体,直至找到具有“卡样”属性的节点实体为止,根据节点实体与“卡样”属性生成图谱查询语句,并在图谱数据库中查询到与商旅航空卡下的不同卡种的具体卡样,作为检索结果。

[0123] 可选的,在根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据之后,还可以包括:如果权重分词中存在目标实体属性的参数值,则根据参数值,对至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,得到与图谱推理引擎对应的检索结果。

[0124] 本实施例中,图谱推理引擎是基于多元数据构建的,可以实现多种图谱推理类型,针对目标问题的权重分词中存在多种并列的目标实体属性,或者存在目标实体属性的参数值,导致无法直接检索到与目标问题对应的答案的情况,图谱推理引擎可以提供相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,以得到检索结果。

[0125] 示例性的,对于目标问题“哪一类信用卡种类最多”,图谱推理引擎先根据“信用卡”、“类型”检索出信用卡的所有类型,然后根据“种类”检索出各类信用卡的种类,最后对

检索出的数据进行推理排序,确定地方特色信用卡种类最多。

[0126] 示例性的,对于目标问题“信用卡的商务系列和经典系列,哪个系列卡种更多”,图谱推理引擎先分别检索出信用卡商务系列和经典系列的卡种数量,然后对两者的卡种数量进行对比计算,确定检索结果为商务系列。对于目标问题“龙卡益贷信用卡和龙卡人民币信用卡是同一类信用卡吗”,图谱推理引擎先分别确定两种卡的上一层实体,然后将两者进行对比,由于两者不一样,因此,确定检索结果为不是。

[0127] 示例性的,对于目标问题“信用卡的商务系列有哪些卡种”,图谱推理引擎检索到商务系列的所有卡种后,对检索出的数据进行列表生成,得到包括商务系列所有卡种的列表。

[0128] 示例性的,对于目标问题“腾讯e龙卡如何开卡”,图谱推理引擎先推理到信用卡节点,才能获取到开卡操作。对于多意图的目标问题“腾讯e龙卡申请和开卡如何办理”,可以将第一个意图如何申请腾讯e龙卡作为直接答案,其余的作为推荐答案。对于目标问题“我是浙江分行开的百大龙卡,请问年费是怎么收取的”,图谱推理引擎可以直接定位到百大龙卡信用卡节点下的年费属性,参数值为浙江分行,进而检索到与目标问题对应的检索结果。

[0129] 步骤240、调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

[0130] 本实施例中,还可以采用通用检索引擎,针对非结构化文档数据,检索到与目标问题对应的答案,作为检索结果。

[0131] 可选的,调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果,可以包括:调用通用检索引擎的接口,根据预设的关联词词库对权重分词进行关联词替换,并为替换后的权重分词设置通用权重;根据权重分词,在预设的索引文件中进行倒排索引,得到索引数据;根据排序模型,对索引数据进行二次排序,得到与通用检索引擎对应的检索结果。

[0132] 本实施例中,如图2b所示,为了能更加准确的找到与目标问题对应的答案,预先建立关联词词库,以在调用通用检索引擎的接口之后,可以对权重分词进行关联词替换。例如,针对目标问题“信用卡丢了怎么办”,意图识别后得到权重分词“丢了”,根据关联词词库可以找到“丢了”的关联词为“挂失”,则可以用“挂失”替换“丢了”,并为“挂失”设置通用权重,得到新的权重分词。然后通用检索引擎根据权重分词对预设的索引文件进行倒排索引,倒排索引实际上是根据属性查找索引数据,例如,根据“华为手机”,找到出现华为手机的非结构化文档。最后,根据排序模型对索引数据进行二次排序,提高数据排序的准确性,得到与通用检索引擎对应的检索结果。

[0133] 可选的,还可以包括:获取数据源的更新数据;对更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;从更新数据中提取实体以及实体属性,并在实体以及实体属性之间建立关联关系;根据分类标签以及关联关系,将更新数据更新到预设的索引文件中。

[0134] 本实施例中,通用检索引擎的检索数据存储在索引文件中,如图2b所示,当数据源有数据更新时,会将更新的数据存储到消息队列中并进行合法性校验,合法性校验是判断更新的数据中是否有指定的词语,如果有,则认为数据合法,如果没有,则认为数据不合法。然后为校验成功的更新数据添加分类标签,例如,操作类、产品类、流程类等,以便于根据标

签进行倒排索引,例如,查询产品类股票相关的文档。然后对更新数据进行预计算,即从更新数据中提取实体以及实体属性,并在两者之间建立关联关系,根据分类标签以及关联关系,将消息队列中校验成功的更新数据更新到索引文件中。

[0135] 其中,可以根据实体和实体属性直接的关联关系对索引文件进行更新,以在检索时缩小检索范围,也可以根据关联关系对排序模型进行更新,以增加排序模型对索引数据进行二次排序的准确性。

[0136] 可选的,获取数据源的更新数据,可以包括:根据数据库归档日志,从与数据源对应的数据库中获取数据源的更新数据;或者通过数据源提供的接口,获取数据源的更新数据;或者根据数据源发送的批量文件,获取数据源的更新数据。

[0137] 本实施例中,可以通过以下几种方式获取数据源的更新数据,如图2b所示,可以通过数据库归档日志,即bin-log日志,结合消息队列实现更新数据实时同步;通过数据源提供的接口实现更新数据实时同步;或者,通过批量文件的方式实现更新数据实时同步。

[0138] 可选的,排序模型包括梯度提升树。

[0139] 本实施例中,排序模型可以是梯度提升树(Gradient Boosting Decision Tree, GBDT树),其中,GBDT树的每个节点都是一个维度,通过使用GBDT树可以对检索结果从多个维度进行二次排序。

[0140] 其中,步骤210、步骤220、步骤230可以同时执行,也可以依次执行,三者之间的执行顺序可调,并在步骤210、步骤220、步骤230都执行完成,得到与各种检索引擎分别对应的检索结果之后,才可以执行步骤250。

[0141] 步骤250、采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0142] 本发明实施例的技术方案,通过获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户,解决了现有技术中存在的检索准确度不高以及二次搜索频繁的问题,实现对多种类型数据进行聚合检索,扩大检索数据的支持范围,提高检索的准确度。

[0143] 实施例三

[0144] 图3是本发明实施例三中的一种基于多元数据的聚合检索装置的结构示意图,本实施例可适用于基于多种类型的数据进行聚合检索的情况,该装置可以由硬件和/或软件来实现,并一般可以集成在提供数据检索服务的电子设备中。如图3所示,该装置包括:

[0145] 获取模块310,用于获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;

[0146] 检索模块320,用于采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;

[0147] 打分模块330,用于采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0148] 本发明实施例的技术方案,通过获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户,解决了现有技术中存在的检索准确度不高以及二次搜索频繁的问题,实现对多种类型数据进行聚合检索,扩大检索数据的支持范围,提高检索的准确度。

[0149] 可选的,获取模块310,用于:获取待检索的目标问题,并对目标问题进行归一化处理,得到初始分词,归一化处理包括:词过滤处理、词映射处理以及设置通用权重;对各初始分词进行句法分析,根据各初始分词的句子结构筛选出有效分词;从有效分词中提取实体以及实体属性,根据实体类别以及实体属性的词性为各有效分词设置图谱权重,得到权重分词。

[0150] 可选的,检索模块320,包括:

[0151] 问答检索单元,用于调用问答引擎的接口,在问答对数据库中检索与权重分词匹配的问题,并将与问题匹配的答案作为检索结果;

[0152] 图谱检索单元,用于调用图谱推理引擎的接口,根据与权重分词对应的图谱查询语句查询图谱数据库,得到检索结果;

[0153] 通用检索单元,用于调用通用检索引擎的接口,根据权重分词对预设的索引文件进行索引,得到检索结果。

[0154] 可选的,图谱检索单元,包括:

[0155] 获取子单元,用于调用图谱推理引擎的接口,获取权重分词中的目标实体以及目标实体属性;

[0156] 查询子单元,用于根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

[0157] 可选的,查询子单元,用于:如果目标实体没有目标实体属性,则定位到目标实体下的具有目标实体属性的节点实体;根据节点实体与目标实体属性生成图谱查询语句,并在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据作为检索结果。

[0158] 可选的,查询子单元,还用于:在根据目标实体以及目标实体属性生成图谱查询语句,在图谱数据库中查询与图谱查询语句匹配的至少一项结构化图谱数据之后,如果权重分词中存在目标实体属性的参数值,则根据参数值,对至少一项结构化图谱数据进行相应的推理排序操作、对比计算操作、列表生成操作或者多层次推理操作,得到与图谱推理引擎对应的检索结果。

[0159] 可选的,通用检索单元,包括:

[0160] 替换子单元,用于调用通用检索引擎的接口,根据预设的关联词词库对权重分词进行关联词替换,并为替换后的权重分词设置通用权重;

[0161] 索引子单元,用于根据权重分词,在预设的索引文件中进行倒排索引,得到索引数据;

[0162] 排序子单元,用于根据排序模型,对索引数据进行二次排序,得到与通用检索引擎对应的检索结果。

[0163] 可选的,更新模块,用于:获取数据源的更新数据;对更新数据进行合法性校验,并对校验成功的更新数据添加分类标签;

[0164] 从更新数据中提取实体以及实体属性,并在实体以及实体属性之间建立关联关系;根据分类标签以及关联关系,将更新数据更新到预设的索引文件中。

[0165] 可选的,更新模块,具体用于:根据数据库归档日志,从与数据源对应的数据库中获取数据源的更新数据;或者通过数据源提供的接口,获取数据源的更新数据;或者根据数据源发送的批量文件,获取数据源的更新数据。

[0166] 可选的,排序模型包括梯度提升树。

[0167] 可选的,还包括:多维度关联模块,用于:在采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户之后,

[0168] 采用半监督学习的方式,从非结构化文档数据中提取目标信息,目标信息包括实体、实体属性以及分类标签;

[0169] 根据目标信息对结构化图谱数据进行归并处理,并对目标信息与对应的原实体、原实体属性进行关联计算,得到更新后的结构化图谱数据;

[0170] 根据更新后的结构化图谱数据,为检索结果提供多维度关联信息。

[0171] 本发明实施例所提供的基于多元数据的聚合检索装置可执行本发明任意实施例所提供的基于多元数据的聚合检索方法,具备执行方法相应的功能模块和有益效果。

[0172] 实施例四

[0173] 图4为本发明实施例公开的一种电子设备的结构示意图。图4示出了适于用来实现本发明实施方式的示例性设备12的框图。图4显示的设备12仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0174] 如图4所示,设备12以通用计算设备的形式表现。设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储器28,连接不同系统组件(包括系统存储器28和处理单元16)的总线18。

[0175] 总线18表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构 (ISA) 总线,微通道体系结构 (MAC) 总线,增强型ISA总线、视频电子标准协会 (VESA) 局域总线以及外围组件互连 (PCI) 总线。

[0176] 设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0177] 系统存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器 (RAM) 30和/或高速缓存存储器32。设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图4未显示,通常称为“硬盘驱动器”)。尽管图4中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM, DVD-ROM或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。存储器28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发明各实施

例的功能。

[0178] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如存储器28中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本发明所描述的实施例中的功能和/或方法。

[0179] 设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信,还可与一个或者多个使得用户能与该设备12交互的设备通信,和/或与使得该设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且,设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器20通过总线18与设备12的其它模块通信。应当明白,尽管图中未示出,可以结合设备12使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0180] 处理单元16通过运行存储在系统存储器28中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的基于多元数据的聚合检索方法。

[0181] 也即:实现一种基于多元数据的聚合检索方法,包括:获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0182] 实施例五

[0183] 本发明实施例五还提供一种计算机可读存储介质,其上存储有计算机程序,程序在被计算机处理器执行时用于执行一种基于多元数据的聚合检索方法,该方法包括:

[0184] 获取待检索的目标问题,并对目标问题进行意图识别处理,得到权重分词;采用多种检索引擎对权重分词进行检索,得到与各检索引擎对应的检索结果,检索引擎包括:与问答对匹配的问答引擎、与结构化图谱数据匹配的图谱推理引擎,以及与非结构化文档数据匹配的通用检索引擎;采用聚合引擎,根据各权重分词的权重,对与各检索引擎对应的检索结果进行置信度打分,并将得分大于预设阈值的检索结果顺序展示给用户。

[0185] 本发明实施例的计算机存储介质,可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是,但不限于,电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0186] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,

其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0187] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0188] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码,程序设计语言包括面向对象的程序设计语言,诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言,诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络,包括局域网(LAN)或广域网(WAN),连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0189] 注意,上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解,本发明不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此,虽然通过以上实施例对本发明进行了较为详细的说明,但是本发明不仅仅限于以上实施例,在不脱离本发明构思的情况下,还可以包括更多其他等效实施例,而本发明的范围由所附的权利要求范围决定。

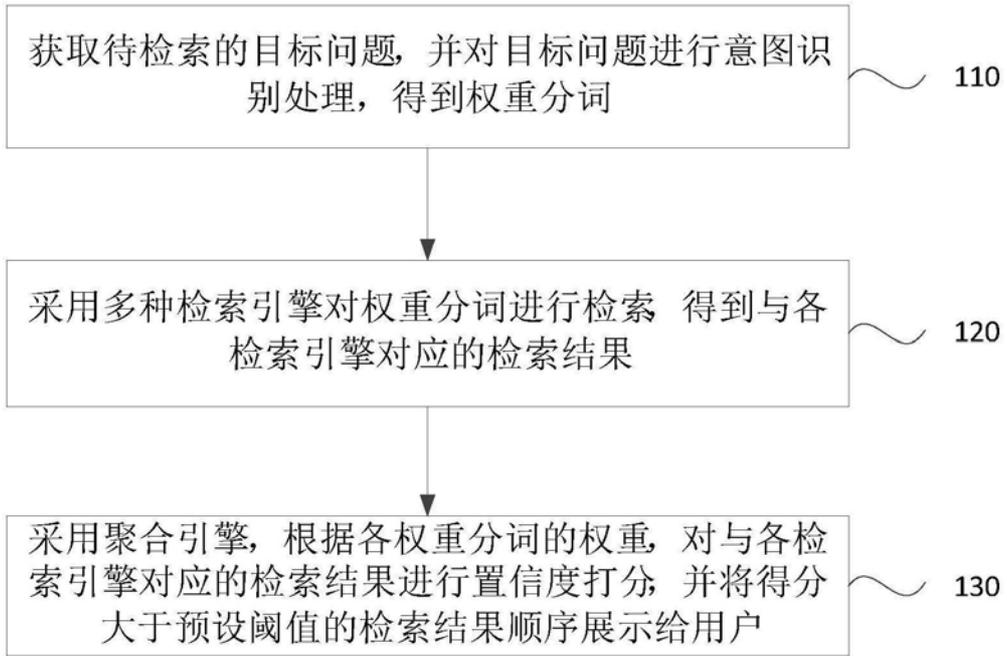


图1a

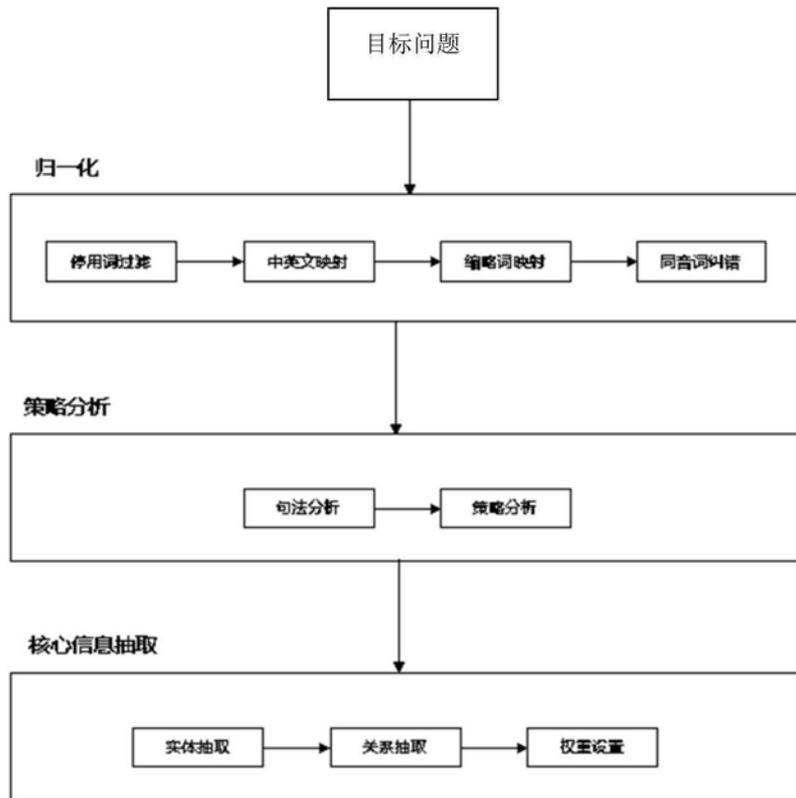


图1b

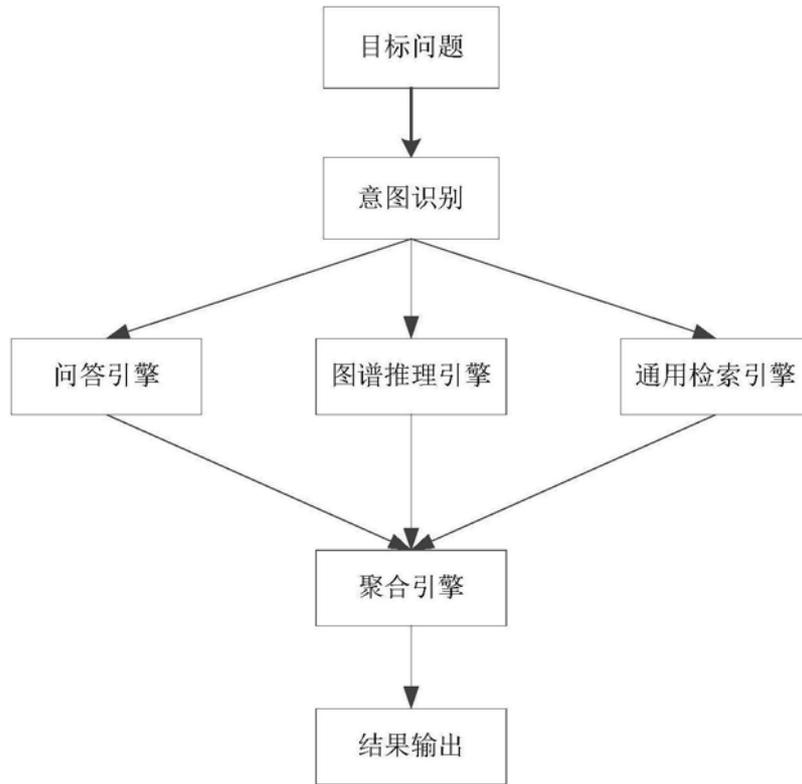


图1c

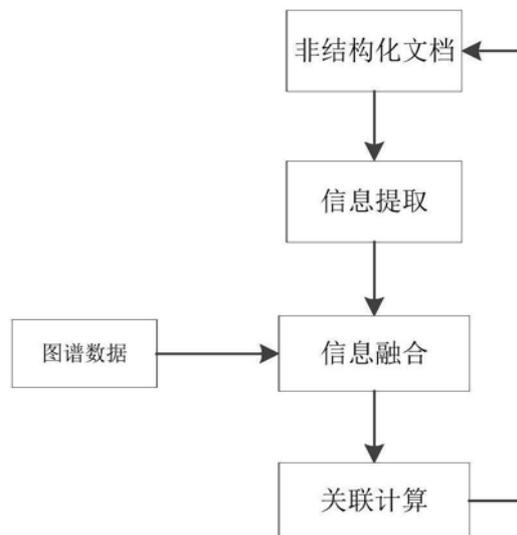


图1d

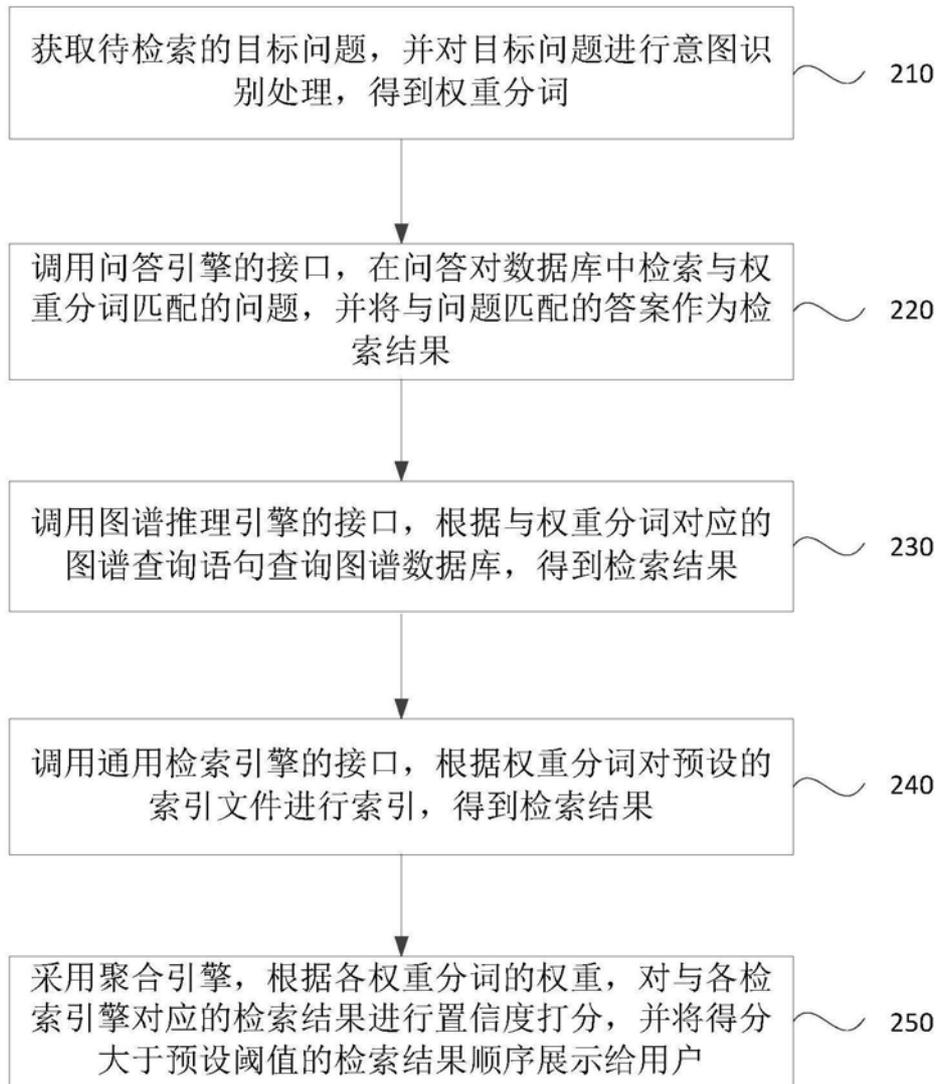


图2a

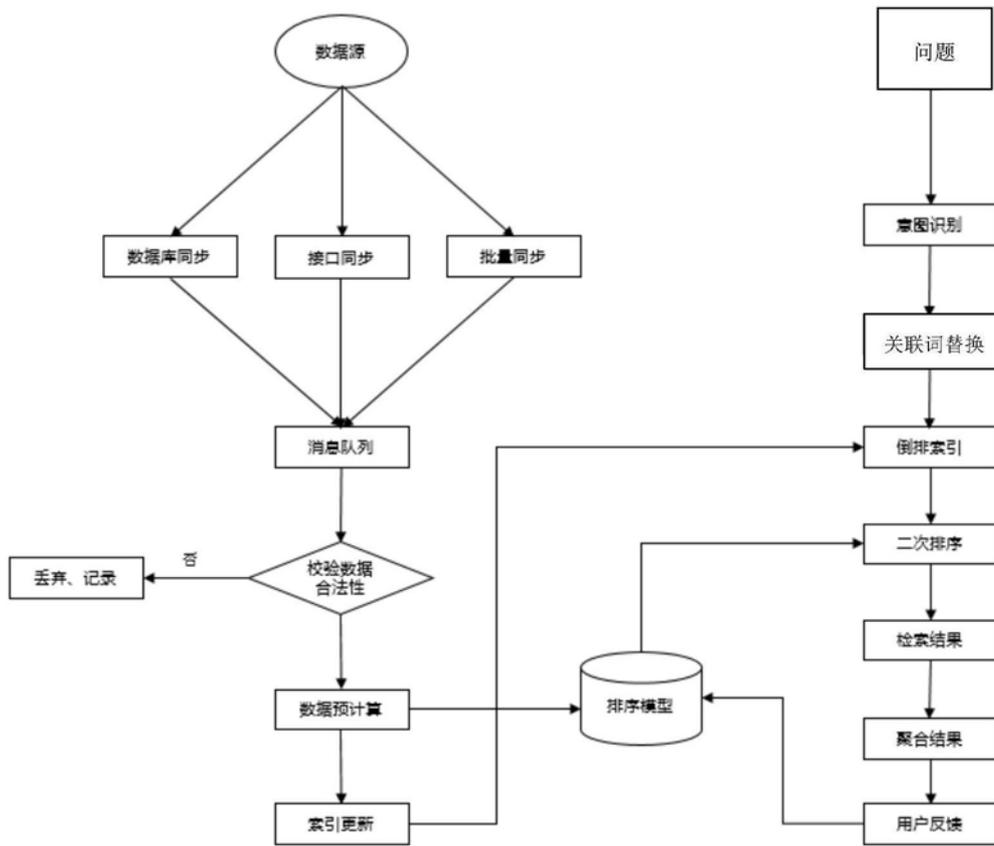


图2b

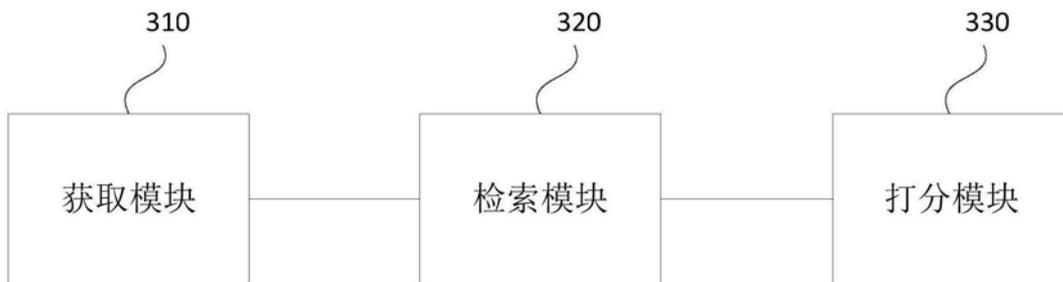


图3

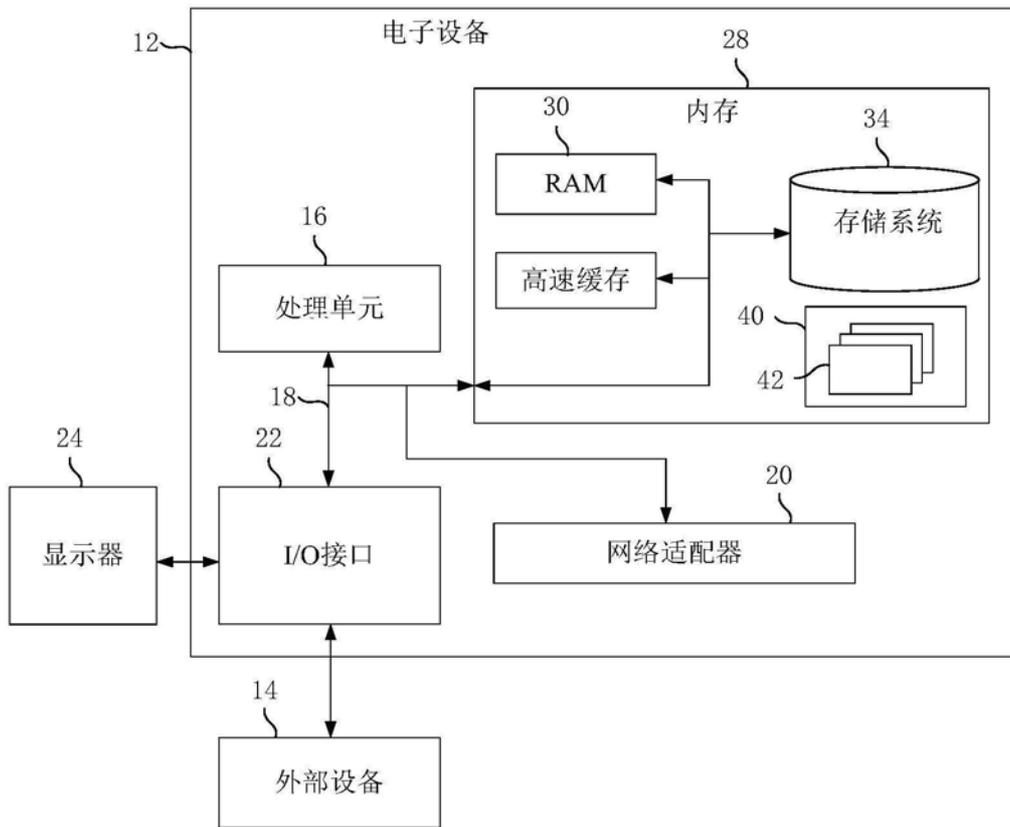


图4