(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 107368468 A
(43)申请公布日 2017.11.21

(21)申请号 201710417415.2

(22)申请日 2017.06.06

(71)申请人 广东广业开元科技有限公司
地址 510623 广东省广州市天河区珠江新城金穗路1号邦华环球广场408

(72)发明人 蔡禹 王晓佳 高峰 孔祥明

(74)专利代理机构 广州嘉权专利商标事务所有限公司 44205
代理人 胡辉

(51)Int.Cl.
G06F 17/27(2006.01)
G06F 17/30(2006.01)
G06N 99/00(2010.01)

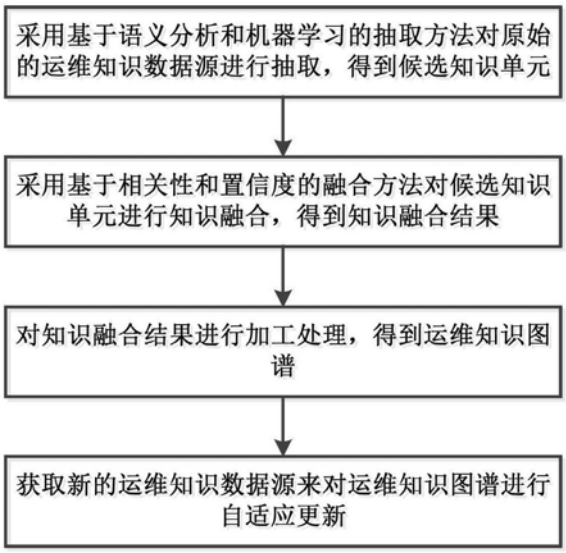
权利要求书4页 说明书13页 附图3页

(54)发明名称

一种运维知识图谱的生成方法及系统

(57)摘要

本发明公开了一种运维知识图谱的生成方法及系统，方法包括：采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果；对知识融合结果进行加工处理，得到运维知识图谱，运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；获取新的运维知识数据源来对运维知识图谱进行自适应更新。本发明包括获取新的运维知识数据源来对运维知识图谱进行自适应更新的步骤，实时性更高；综合采用了基于语义分析和机器学习的抽取方法和基于相关性和置信度的融合方法，效率更高，成本更低且更加方便。本发明可广泛应用于计算机应用领域。



CN 107368468 A

1.一种运维知识图谱的生成方法,其特征在于:包括以下步骤:

采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取,得到候选知识单元;

采用基于相关性和置信度的融合方法对候选知识单元进行知识融合,得到知识融合结果;

对知识融合结果进行加工处理,得到运维知识图谱,所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成;

获取新的运维知识数据源来对运维知识图谱进行自适应更新。

2.根据权利要求1所述的一种运维知识图谱的生成方法,其特征在于:所述采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取,得到候选知识单元这一步骤,其包括:

对运维信息化系统进行信息自动采集,得到原始的运维知识数据源;

采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取,得到候选知识单元。

3.根据权利要求2所述的一种运维知识图谱的生成方法,其特征在于:所述对运维信息化系统进行信息自动采集,得到原始的运维知识数据源这一步骤,其具体为:

采用分布式爬虫和接口对运维信息化系统中的运维工单或系统日志进行信息抽取,得到原始的运维知识数据源。

4.根据权利要求2所述的一种运维知识图谱的生成方法,其特征在于:所述采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取,得到候选知识单元这一步骤,其包括:

通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析,形成语法树并找出每个语句的名词短语;

通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系,从而形成由名词短语对和名词短语间的相关关系组成的三元组;

以所有三元组中的任一三元组作为当前三元组,判断当前三元组是否满足设定的候选条件,若是,则将当前三元组标记为候选抽取的三元组,反之,则对当前三元组进行归档暂不处理操作;

采用朴素贝叶斯分类器判断候选抽取的三元组是否可信,若是,则将该候选抽取的三元组抽取出来作为可信的三元组,反之,则对该候选抽取的三元组进行归档暂不处理操作;

对可信的三元组进行存储和归并,从而得到由最终的抽取结果组成的候选知识单元,所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

5.根据权利要求4所述的一种运维知识图谱的生成方法,其特征在于:所述采用基于相关性和置信度的融合方法对候选知识单元进行知识融合,得到知识融合结果这一步骤,其包括:

以候选知识单元作为输入文本,对输入文本进行实体识别并生成候选实体;

对生成的候选实体进行实体相关性计算并构造相应的实体相关图,其中,实体相关图的顶点采用实体对象构造而成,实体相关图的边通过语言相关性权重计算后确定,所述语言相关性权重计算的公式为:

$$\begin{cases} w_{ij}=\begin{cases} 1 & e_i和e_j存在直接关系 \\ GD(e_i,e_j) & e_i和e_j存在间接关系 \end{cases} \\ GD(e_i,e_j)=1-\dfrac{\log(\max(|in(e_i)|,|in(e_j)|))-\log(|in(e_i)|\cap|in(e_j)|)}{\log(|Kb|)-\log(\min(|in(e_i)|,|in(e_j)|))} \end{cases},$$

其中，e_i和e_j分别为候选实体中名词短语对(e_i,e_j)的2个名词短语，且i<j，w_{ij}为e_i和e_j的语言相关性权重，$GD(e_i,e_j)$为e_i和e_j存在间接关系时的语言相关性权重，$in(e_i)$和$in(e_j)$分别表示知识库中与e_i和e_j所表示的候选实体存在指向关系的实体集合，log、max、min和∩分别为对数函数符号、取最大值运算符号、取最小值运算符号和交集符号，Kb为实体相关图的所有实体集合，|Kb|表示集合Kb中的元素个数；

根据构造的实体相关图进行候选实体顶点的置信度计算，得到候选实体顶点的置信度分数，所述候选实体顶点的置信度分数计算公式为：

$$PR(v_a)=(1-\alpha)docSim(v_a)+\alpha\sum_{v_b\in Nh(v_a)}\frac{w_{ba}}{\sum_{v_k\in Nh(v_b)}w_{bk}}PR(v_b),$$

其中，v_a为候选实体顶点，v_b和v_k均为实体相关图的实体顶点，$Nh(v_a)$和$Nh(v_b)$分别为顶点v_a和顶点v_b的邻域，α为阻尼因子，$PR(v_a)$和$PR(v_b)$分别为v_a和v_b的置信度分数，$docSim(v_a)$为顶点v_a所表示的候选实体与输入文本的上下文相似度，w_{ba}为实体相关图中边(v_a,v_b)的权重，w_{bk}为实体相关图中边(v_b,v_k)的权重；

进行候选实体与输入文本的语义相关性计算，所述候选实体顶点v_a与输入文本D的语义相关性$SR(v_a,D)$计算公式为：

$$SR(v_a,D)=\sum_{v_{k0}\in N_{\max R}}w_{ak0}\cdot PR(v_{k0}),$$

其中，v_{k0}为实体顶点，N_{maxR}为输入文本D中的每个实体指称项对应的候选集合中相关度最高的候选实体构成的子集，w_{ak0}为实体相关图中边(v_a,v_{k0})的权重，$PR(v_{k0})$为v_{k0}的置信度分数；

根据置信度计算的结果和语义相关性计算的结果进行语义一致性计算，并根据语义一致性计算的结果得到知识融合结果，所述候选实体m与实体指称项c_{k0}的语义一致性$SCC(m,c_{k0})$计算公式为：

$$SCC(m,c_{k0})=\frac{PR(v_{k0})+SR(v_{k0},D)}{\sum_{V_j\in V_{k0}}PR(v_j)+SR(v_j,D)}。$$

6.根据权利要求1-5任一项所述的一种运维知识图谱的生成方法，其特征在于：所述对知识融合结果进行加工处理，得到运维知识图谱这一步骤，其包括：

对知识融合结果进行实体并列关系相似度计算，得到运维知识实体间的并列关系相似度；

对知识融合结果进行实体上下级关系抽取，从而确定运维知识实体的上下级关系；

对确定的所有运维知识实体上下级关系进行聚类，并对聚类的结果进行语义类的标定，从而生成运维知识实体的本体；

从已有的运维知识实体关系数据出发，根据运维知识实体间的并列关系相似度和运维知识实体的本体进行知识推理，得到运维知识实体间的新关联和对应的运维知识图谱。

7.根据权利要求2-5任一项所述的一种运维知识图谱的生成方法，其特征在于：所述获取新的运维知识数据源来对运维知识图谱进行自适应更新这一步骤，其包括：

通过对运维信息化系统进行信息自动采集实时获取新的运维知识数据源；

对新的运维知识数据源进行预处理，并将预处理后的运维知识数据源中的数据分别标记为第一数据和第二数据，所述第一数据是指与现有运维知识图谱的数据的差异大于设定的差异阈值的数据，所述第二数据是指与现有运维知识图谱的数据的差异小于等于设定的差异阈值的数据；

以第一数据作为原始的运维知识数据源，返回采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元这一步骤，最终得到第一数据对应的运维知识图谱，并将第一数据对应的运维知识图谱补充到运维知识图谱数据库中；

分析出第二数据中区别于现有运维知识图谱的运维知识实体和第一运维知识实体关系，并判断第二数据的时序性是否小于1，若是，则将第二数据剔除，反之，则将第一运维知识实体关系标记为历史数据，然后对第一运维知识实体关系的时序性进行计算和排序，并根据计算和排序的结果更新现有运维知识图谱。

8.一种运维知识图谱的生成系统，其特征在于：包括：

知识抽取模块，用于采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；

知识融合模块，用于采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果；

知识加工模块，用于对知识融合结果进行加工处理，得到运维知识图谱，所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；

知识更新模块，用于获取新的运维知识数据源来对运维知识图谱进行自适应更新。

9.根据权利要求8所述的一种运维知识图谱的生成系统，其特征在于：所述知识抽取模块包括：

信息自动采集单元，用于对运维信息化系统进行信息自动采集，得到原始的运维知识数据源；

信息抽取单元，用于采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元。

10.根据权利要求9所述的一种运维知识图谱的生成系统，其特征在于：所述信息抽取单元包括：

语法分析子单元，用于通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析，形成语法树并找出每个语句的名词短语；

三元组构建子单元，用于通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系，从而形成由名词短语对和名词短语间的相关关系组成的三元组；

候选抽取三元组判断子单元，用于以所有三元组中的任一三元组作为当前三元组，判断当前三元组是否满足设定的候选条件，若是，则将当前三元组标记为候选抽取的三元组，反之，则对当前三元组进行归档暂不处理操作；

可信判断子单元，用于采用朴素贝叶斯分类器判断候选抽取的三元组是否可信，若是，则将该候选抽取的三元组抽取出来作为可信的三元组，反之，则对该候选抽取的三元组进行归档暂不处理操作；

存储归并子单元，用于对可信的三元组进行存储和归并，从而得到由最终的抽取结果组成的候选知识单元，所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

一种运维知识图谱的生成方法及系统

技术领域

[0001] 本发明涉及计算机应用领域，尤其是一种运维知识图谱的生成方法及系统。

背景技术

[0002] 在知识经济到来的今天，知识已被企业提升到战略资源的位置，企业采用知识管理势在必行。随着业务系统越来越庞大、业务逻辑越来越复杂、系统变更越来越频繁、工作要求越来越高，在业务支撑网运营管理工作的难度也越来越大。通过建设内容丰富和人人参与的统一知识库，可达到为企业建立知识上传和下达的渠道、打造学习型业务支撑团队、助力公司长期可持续健康发展的目的。

[0003] 知识管理(KM,Knowledge Management)是网络新经济时代的新兴管理思潮与方法，管理学者彼得.杜拉克早在一九六五年即预言：“知识将取代土地、劳动、资本与机器设备，成为最重要的生产因素。”受到20世纪90年代的信息化(资讯化)蓬勃发展影响，知识管理的观念结合网际网络构建的入口网站、数据库以及应用电脑软件系统等工具，成为累积知识财富，创造更多竞争力的新世纪利器。

[0004] 而知识图谱就是一个很好的知识管理手段。自语义网的概念提出，语义Web数据源的数量激增，互联网正从仅包含网页和网页之间超链接的文档万维网转变成包含大量描述各种实体和实体之间丰富关系的数据万维网。在此背景下，知识图谱于2012年5月首先由Google公司提出，其目标在于描述各种实体与概念，及实体、概念之间的关联关系，从而改善搜索结果。紧随其后，搜狗、微软、百度等公司相继提出各自的知识图谱产品。

[0005] 然而现有的知识图谱构建方法，大多无法时更新已有的知识图谱，实时性较低，难以满足实时性要求高的应用场合要求。

[0006] 随着IT技术的不断发展，运维信息化得到了越来越多人的重视。然而，受数据源不足、使用场景不明等因素的影响，知识图谱一直未能被应用于运维信息化领域。目前运维信息化领域仍依靠人工录入信息的方式来进行知识的积累，效率低，成本高，且不能描述知识之间的关系，不够方便，亟待进一步完善和提高。

发明内容

[0007] 为解决上述技术问题，本发明的目的在于：提供一种实时、效率高、成本低和方便的，运维知识图谱的生成方法。

[0008] 本发明的另一目的在于：提供一种实时、效率高、成本低和方便的，运维知识图谱的生成系统。

[0009] 本发明所采取的技术方案是：

[0010] 一种运维知识图谱的生成方法，包括以下步骤：

[0011] 采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；

[0012] 采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融

合结果；

[0013] 对知识融合结果进行加工处理，得到运维知识图谱，所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；

[0014] 获取新的运维知识数据源来对运维知识图谱进行自适应更新。

[0015] 进一步，所述采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元这一步骤，其包括：

[0016] 对运维信息化系统进行信息自动采集，得到原始的运维知识数据源；

[0017] 采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元。

[0018] 进一步，所述对运维信息化系统进行信息自动采集，得到原始的运维知识数据源这一步骤，其具体为：

[0019] 采用分布式爬虫和接口对运维信息化系统中的运维工单或系统日志进行信息抽取，得到原始的运维知识数据源。

[0020] 进一步，所述采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元这一步骤，其包括：

[0021] 通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析，形成语法树并找出每个语句的名词短语；

[0022] 通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系，从而形成由名词短语对和名词短语间的相关关系组成的三元组；

[0023] 以所有三元组中的任一三元组作为当前三元组，判断当前三元组是否满足设定的候选条件，若是，则将当前三元组标记为候选抽取的三元组，反之，则对当前三元组进行归档暂不处理操作；

[0024] 采用朴素贝叶斯分类器判断候选抽取的三元组是否可信，若是，则将该候选抽取的三元组抽取出来作为可信的三元组，反之，则对该候选抽取的三元组进行归档暂不处理操作；

[0025] 对可信的三元组进行存储和归并，从而得到由最终的抽取结果组成的候选知识单元，所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

[0026] 进一步，所述采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果这一步骤，其包括：

[0027] 以候选知识单元作为输入文本，对输入文本进行实体识别并生成候选实体；

[0028] 对生成的候选实体进行实体相关性计算并构造相应的实体相关图，其中，实体相关图的顶点采用实体对象构造而成，实体相关图的边通过语言相关性权重计算后确定，所述语言相关性权重计算的公式为：

[0029]
$$\begin{cases} w_{ij}=\begin{cases} 1 & e_i\text{和}e_j\text{存在直接关系} \\ GD(e_i,e_j) & e_i\text{和}e_j\text{存在间接关系} \end{cases} \\ GD(e_i,e_j)=1-\dfrac{\log(\max(|in(e_i)|,|in(e_j)|))-\log(|in(e_i)|\cap|in(e_j)|)}{\log(|Kb|)-\log(\min(|in(e_i)|,|in(e_j)|))} \end{cases},$$

[0030] 其中，e_i和e_j分别为候选实体中名词短语对(e_i,e_j)的2个名词短语，且$i<j$，w_{ij}为e_i和e_j的语言相关性权重，$GD(e_i,e_j)$为e_i和e_j存在间接关系时的语言相关性权重，$in(e_i)$和$in(e_j)$分别表示知识库中与e_i和e_j所表示的候选实体存在指向关系的实体集合，log、max、min和∩分别为对数函数符号、取最大值运算符号、取最小值运算符号和交集符号，Kb为实体相关图的所有实体集合，|Kb|表示集合Kb中的元素个数；

[0031] 根据构造的实体相关图进行候选实体顶点的置信度计算，得到候选实体顶点的置信度分数，所述候选实体顶点的置信度分数计算公式为：

[0032] $$PR(v_a)=(1-\alpha)docSim(v_a)+\alpha\sum_{v_b\in Nh(v_a)}\frac{w_{ba}}{\sum_{v_k\in Nh(v_b)}w_{bk}}PR(v_b),$$

[0033] 其中，v_a为候选实体顶点，v_b和v_k均为实体相关图的实体顶点，$Nh(v_a)$和$Nh(v_b)$分别为顶点v_a和顶点v_b的邻域，α为阻尼因子，$PR(v_a)$和$PR(v_b)$分别为v_a和v_b的置信度分数，$docSim(v_a)$为顶点v_a所表示的候选实体与输入文本的上下文相似度，w_{ba}为实体相关图中边(v_a,v_b)的权重，w_{bk}为实体相关图中边(v_b,v_k)的权重；

[0034] 进行候选实体与输入文本的语义相关性计算，所述候选实体顶点v_a与输入文本D的语义相关性$SR(v_a,D)$计算公式为：

[0035] $$SR(v_a,D)=\sum_{v_{k0}\in N_{\max R}}w_{ak0}\cdot PR(v_{k0}),$$

[0036] 其中，v_{k0}为实体顶点，N_{maxR}为输入文本D中的每个实体指称项对应的候选集合中相关度最高的候选实体构成的子集，w_{ak0}为实体相关图中边(v_a,v_{k0})的权重，$PR(v_{k0})$为v_{k0}的置信度分数；

[0037] 根据置信度计算的结果和语义相关性计算的结果进行语义一致性计算，并根据语义一致性计算的结果得到知识融合结果，所述候选实体m与实体指称项c_{k0}的语义一致性$SCC(m,c_{k0})$计算公式为：

[0038] $$SCC(m,c_{k0})=\frac{PR(v_{k0})+SR(v_{k0},D)}{\sum_{V_j\in V_{k0}}PR(v_j)+SR(v_j,D)}。$$

[0039] 进一步，所述对知识融合结果进行加工处理，得到运维知识图谱这一步骤，其包括：

[0040] 对知识融合结果进行实体并列关系相似度计算，得到运维知识实体间的并列关系相似度；

[0041] 对知识融合结果进行实体上下级关系抽取，从而确定运维知识实体的上下级关系；

[0042] 对确定的所有运维知识实体上下级关系进行聚类，并对聚类的结果进行语义类的标定，从而生成运维知识实体的本体；

[0043] 从已有的运维知识实体关系数据出发，根据运维知识实体间的并列关系相似度和运维知识实体的本体进行知识推理，得到运维知识实体间的新关联和对应的运维知识图

谱。

[0044] 进一步，所述获取新的运维知识数据源来对运维知识图谱进行自适应更新这一步骤，其包括：

[0045] 通过对运维信息化系统进行信息自动采集实时获取新的运维知识数据源；

[0046] 对新的运维知识数据源进行预处理，并将预处理后的运维知识数据源中的数据分别标记为第一数据和第二数据，所述第一数据是指与现有运维知识图谱的数据的差异大于设定的差异阈值的数据，所述第二数据是指与现有运维知识图谱的数据的差异小于等于设定的差异阈值的数据；

[0047] 以第一数据作为原始的运维知识数据源，返回采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元这一步骤，最终得到第一数据对应的运维知识图谱，并将第一数据对应的运维知识图谱补充到运维知识图谱数据库中；

[0048] 分析出第二数据中区别于现有运维知识图谱的运维知识实体和第一运维知识实体关系，并判断第二数据的时序性是否小于1，若是，则将第二数据剔除，反之，则将第一运维知识实体关系标记为历史数据，然后对第一运维知识实体关系的时序性进行计算和排序，并根据计算和排序的结果更新现有运维知识图谱。

[0049] 本发明所采取的另一技术方案是：

[0050] 一种运维知识图谱的生成系统，包括：

[0051] 知识抽取模块，用于采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；

[0052] 知识融合模块，用于采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果；

[0053] 知识加工模块，用于对知识融合结果进行加工处理，得到运维知识图谱，所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；

[0054] 知识更新模块，用于获取新的运维知识数据源来对运维知识图谱进行自适应更新。

[0055] 进一步，所述知识抽取模块包括：

[0056] 信息自动采集单元，用于对运维信息化系统进行信息自动采集，得到原始的运维知识数据源；

[0057] 信息抽取单元，用于采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元。

[0058] 进一步，所述信息抽取单元包括：

[0059] 语法分析子单元，用于通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析，形成语法树并找出每个语句的名词短语；

[0060] 三元组构建子单元，用于通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系，从而形成由名词短语对和名词短语间的相关关系组成的三元组；

[0061] 候选抽取三元组判断子单元，用于以所有三元组中的任一三元组作为当前三元组，判断当前三元组是否满足设定的候选条件，若是，则将当前三元组标记为候选抽取的三元组，反之，则对当前三元组进行归档暂不处理操作；

[0062] 可信判断子单元，用于采用朴素贝叶斯分类器判断候选抽取的三元组是否可信，若是，则将该候选抽取的三元组抽取出来作为可信的三元组，反之，则对该候选抽取的三元组进行归档暂不处理操作；

[0063] 存储归并子单元，用于对可信的三元组进行存储和归并，从而得到由最终的抽取结果组成的候选知识单元，所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

[0064] 本发明的方法的有益效果是：包括获取新的运维知识数据源来对运维知识图谱进行自适应更新的步骤，能获取新的运维知识数据来对已有的对运维知识图谱进行实时更新，实时性更高；综合采用了基于语义分析和机器学习的抽取方法和基于相关性和置信度的融合方法，依次通过抽取、知识融合和加工处理来得出运维知识图谱，基于语义分析和图论的模型来构建运维知识图谱，并通过运维知识图谱来描述运维知识之间的关系，解决了现有技术无法在运维信息化领域应用知识图谱的问题，不再需要依靠人工录入信息的方式来进行知识的积累，效率更高，成本更低且更加方便。

[0065] 本发明的系统的有益效果是：包括用于获取新的运维知识数据源来对运维知识图谱进行自适应更新的知识更新模块，能获取新的运维知识数据来对已有的对运维知识图谱进行实时更新，实时性更高；综合采用了基于语义分析和机器学习的抽取方法和基于相关性和置信度的融合方法，依次执行知识抽取模块、知识融合模块和知识加工模块的操作来得出运维知识图谱，基于语义分析和图论的模型来构建运维知识图谱，并通过运维知识图谱来描述运维知识之间的关系，解决了现有技术无法在运维信息化领域应用知识图谱的问题，不再需要依靠人工录入信息的方式来进行知识的积累，效率更高，成本更低且更加方便。

附图说明

[0066] 图1为本发明一种运维知识图谱的生成方法的整体流程图；

[0067] 图2为本发明实施例一运维知识图谱的构建过程流程图；

[0068] 图3为图2中知识提取/抽取过程的具体流程图；

[0069] 图4为图2中知识融合过程的具体流程图；

[0070] 图5为图2中知识加工过程的具体流程图。

具体实施方式

[0071] 参照图1，一种运维知识图谱的生成方法，包括以下步骤：

[0072] 采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；

[0073] 采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果；

[0074] 对知识融合结果进行加工处理，得到运维知识图谱，所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；

[0075] 获取新的运维知识数据源来对运维知识图谱进行自适应更新。

[0076] 进一步，所述采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源

进行抽取，得到候选知识单元这一步骤，其包括：

[0077] 对运维信息化系统进行信息自动采集，得到原始的运维知识数据源；

[0078] 采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元。

[0079] 进一步作为优选的实施方式，所述对运维信息化系统进行信息自动采集，得到原始的运维知识数据源这一步骤，其具体为：

[0080] 采用分布式爬虫和接口对运维信息化系统中的运维工单或系统日志进行信息抽取，得到原始的运维知识数据源。

[0081] 进一步作为优选的实施方式，所述采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元这一步骤，其包括：

[0082] 通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析，形成语法树并找出每个语句的名词短语；

[0083] 通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系，从而形成由名词短语对和名词短语间的相关关系组成的三元组；

[0084] 以所有三元组中的任一三元组作为当前三元组，判断当前三元组是否满足设定的候选条件，若是，则将当前三元组标记为候选抽取的三元组，反之，则对当前三元组进行归档暂不处理操作；

[0085] 采用朴素贝叶斯分类器判断候选抽取的三元组是否可信，若是，则将该候选抽取的三元组抽取出来作为可信的三元组，反之，则对该候选抽取的三元组进行归档暂不处理操作；

[0086] 对可信的三元组进行存储和归并，从而得到由最终的抽取结果组成的候选知识单元，所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

[0087] 进一步作为优选的实施方式，所述采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果这一步骤，其包括：

[0088] 以候选知识单元作为输入文本，对输入文本进行实体识别并生成候选实体；

[0089] 对生成的候选实体进行实体相关性计算并构造相应的实体相关图，其中，实体相关图的顶点采用实体对象构造而成，实体相关图的边通过语言相关性权重计算后确定，所述语言相关性权重计算的公式为：

[0090]
$$\begin{cases} w_{ij}=\begin{cases} 1 & e_i\text{和}e_j\text{存在直接关系} \\ GD(e_i,e_j) & e_i\text{和}e_j\text{存在间接关系} \end{cases} \\ GD(e_i,e_j)=1-\dfrac{\log(\max(|in(e_i)|,|in(e_j)|))-\log(|in(e_i)|\cap|in(e_j)|)}{\log(|Kb|)-\log(\min(|in(e_i)|,|in(e_j)|))} \end{cases},$$

[0091] 其中，e_i和e_j分别为候选实体中名词短语对(e_i,e_j)的2个名词短语，且$i<j$，w_{ij}为e_i和e_j的语言相关性权重，$GD(e_i,e_j)$为e_i和e_j存在间接关系时的语言相关性权重，$in(e_i)$和$in(e_j)$分别表示知识库中与e_i和e_j所表示的候选实体存在指向关系的实体集合，log、max、min、∩和||分别为对数函数符号、取最大值运算符号、取最小值运算符号、交集符号和求集合中元素个数符号，Kb为实体相关图的所有实体集合，|Kb|表示集合Kb中的元素个数；

[0092] 根据构造的实体相关图进行候选实体顶点的置信度计算，得到候选实体顶点的置信度分数，所述候选实体顶点的置信度分数计算公式为：

[0093] $$PR(v_a)=(1-\alpha)docSim(v_a)+\alpha\sum_{v_b\in Nh(v_a)}\frac{w_{ba}}{\sum_{v_k\in Nh(v_b)}w_{bk}}PR(v_b),$$

[0094] 其中，v_a为候选实体顶点，v_b和v_k均为实体相关图的实体顶点，$Nh(v_a)$和$Nh(v_b)$分别为顶点v_a和顶点v_b的邻域，α为阻尼因子，$PR(v_a)$和$PR(v_b)$分别为v_a和v_b的置信度分数，$docSim(v_a)$为顶点v_a所表示的候选实体与输入文本的上下文相似度，w_{ba}为实体相关图中边(v_a,v_b)的权重，w_{bk}为实体相关图中边(v_b,v_k)的权重；

[0095] 进行候选实体与输入文本的语义相关性计算，所述候选实体顶点v_a与输入文本D的语义相关性$SR(v_a,D)$计算公式为：

[0096] $$SR(v_a,D)=\sum_{v_{k0}\in N_{\max R}}w_{ak0}\cdot PR(v_{k0}),$$

[0097] 其中，v_{k0}为实体顶点，N_{maxR}为输入文本D中的每个实体指称项对应的候选集合中相关度最高的候选实体构成的子集，w_{ak0}为实体相关图中边(v_a,v_{k0})的权重，$PR(v_{k0})$为v_{k0}的置信度分数；

[0098] 根据置信度计算的结果和语义相关性计算的结果进行语义一致性计算，并根据语义一致性计算的结果得到知识融合结果，所述候选实体m与实体指称项c_{k0}的语义一致性$SCC(m,c_{k0})$计算公式为：

[0099] $$SCC(m,c_{k0})=\frac{PR(v_{k0})+SR(v_{k0},D)}{\sum_{V_j\in V_{k0}}PR(v_j)+SR(v_j,D)}。$$

[0100] 在实体相关图中，名词短语e_i和e_j对应实体相关图的顶点，候选实体中名词短语对(e_i,e_j)对应实体相关图中连接顶点e_i和e_j所构成的边。

[0101] $Nh(v_a)$和$Nh(v_b)$分别为顶点v_a和顶点v_b的邻域，即$Nh(v_a)$和$Nh(v_b)$分别表示与顶点v_a和顶点v_b相邻的顶点集合。

[0102] 进一步作为优选的实施方式，所述对知识融合结果进行加工处理，得到运维知识图谱这一步骤，其包括：

[0103] 对知识融合结果进行实体并列关系相似度计算，得到运维知识实体间的并列关系相似度；

[0104] 对知识融合结果进行实体上下级关系抽取，从而确定运维知识实体的上下级关系；

[0105] 对确定的所有运维知识实体上下级关系进行聚类，并对聚类的结果进行语义类的标定，从而生成运维知识实体的本体；

[0106] 从已有的运维知识实体关系数据出发，根据运维知识实体间的并列关系相似度和运维知识实体的本体进行知识推理，得到运维知识实体间的新关联和对应的运维知识图谱。

[0107] 进一步作为优选的实施方式，所述获取新的运维知识数据源来对运维知识图谱进

行自适应更新这一步骤，其包括：

[0108] 通过对运维信息化系统进行信息自动采集实时获取新的运维知识数据源；

[0109] 对新的运维知识数据源进行预处理，并将预处理后的运维知识数据源中的数据分别标记为第一数据和第二数据，所述第一数据是指与现有运维知识图谱的数据的差异大于设定的差异阈值的数据，所述第二数据是指与现有运维知识图谱的数据的差异小于等于设定的差异阈值的数据；

[0110] 以第一数据作为原始的运维知识数据源，返回采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元这一步骤，最终得到第一数据对应的运维知识图谱，并将第一数据对应的运维知识图谱补充到运维知识图谱数据库中；

[0111] 分析出第二数据中区别于现有运维知识图谱的运维知识实体和第一运维知识实体关系，并判断第二数据的时序性是否小于1，若是，则将第二数据剔除，反之，则将第一运维知识实体关系标记为历史数据，然后对第一运维知识实体关系的时序性进行计算和排序，并根据计算和排序的结果更新现有运维知识图谱。

[0112] 其中，运维知识图谱数据库用于存储运维知识图谱。

[0113] 本发明一种运维知识图谱的生成系统，包括：

[0114] 知识抽取模块，用于采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元；

[0115] 知识融合模块，用于采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果；

[0116] 知识加工模块，用于对知识融合结果进行加工处理，得到运维知识图谱，所述运维知识图谱由运维知识实体和运维知识实体间的相互关系组成；

[0117] 知识更新模块，用于获取新的运维知识数据源来对运维知识图谱进行自适应更新。

[0118] 进一步作为优选的实施方式，所述知识抽取模块包括：

[0119] 信息自动采集单元，用于对运维信息化系统进行信息自动采集，得到原始的运维知识数据源；

[0120] 信息抽取单元，用于采用自然语言分析器和分类器对原始的运维知识数据源进行信息抽取，得到候选知识单元。

[0121] 进一步作为优选的实施方式，所述信息抽取单元包括：

[0122] 语法分析子单元，用于通过自然语言分析器对原始的运维知识数据源中所有语句进行语法分析，形成语法树并找出每个语句的名词短语；

[0123] 三元组构建子单元，用于通过语法树构建每个语句中所有的名词短语对和每对名词短语间的相关关系，从而形成由名词短语对和名词短语间的相关关系组成的三元组；

[0124] 候选抽取三元组判断子单元，用于以所有三元组中的任一三元组作为当前三元组，判断当前三元组是否满足设定的候选条件，若是，则将当前三元组标记为候选抽取的三元组，反之，则对当前三元组进行归档暂不处理操作；

[0125] 可信判断子单元，用于采用朴素贝叶斯分类器判断候选抽取的三元组是否可信，若是，则将该候选抽取的三元组抽取出来作为可信的三元组，反之，则对该候选抽取的三元

组进行归档暂不处理操作；

[0126] 存储归并子单元，用于对可信的三元组进行存储和归并，从而得到由最终的抽取结果组成的候选知识单元，所述最终的抽取结果只存储各个不同的三元组及各个不同的三元组出现的频次。

[0127] 下面结合说明书附图和具体实施例对本发明作进一步解释和说明。

[0128] 实施例一

[0129] 参照图2-5，本发明的第一实施例：

[0130] 针对现有技术无法实时更新已有的知识图谱以及无法将知识图谱应用于运维信息化领域的问题，本发明提出了一种新的运维知识图谱的生成技术。该生成技术首先通过爬虫等方式对运维知识数据源进行信息的提取或抽取，然后进行知识融合、知识验证、知识计算、知识存储等一系列流程来构建运维知识图谱，并能在知识融合、知识验证和知识计算过程中，使得运维知识图谱可以形成并不断进行丰富和自我修正，最终可以得到一个高质量的运维知识库。

[0131] 下面从名词解释和具体实现过程以及实现原理这两方面入手对本发明的运维知识图谱生成技术进行详细说明。

[0132] (一)名词解释

[0133] 本发明涉及到的专有名词如下：

[0134] 深度学习：源于人工神经网络的研究，通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。

[0135] 知识库：知识工程中结构化、易操作、易利用和全面有组织的知识集群，是针对某一(或某些)领域问题求解的需要，采用某种(或若干)知识表示方式在计算机存储器中存储、组织、管理和使用的相互联系的知识片集合。这些知识片包括与领域相关的理论知识，事实数据，由专家经验得到的启发式知识(如某领域内有关的定义、定理和运算法则等)，以及常识性知识等。一般的应用程序与基于知识的系统之间的区别在于：一般的应用程序是把问题求解的知识隐含地编码在程序中，而基于知识的系统则将应用领域的问题求解知识显式地表达，并单独地组成一个相对独立的程序实体。

[0136] 运维信息化系统：以IT部门在日常的运行维护管理流程为核心，以事件跟踪为主线，以解决IT运维管理中的八大管理问题(流程管理、事件管理、问题管理、变更管理、发布管理、运行管理、知识管理、综合分析管理)为目的，为IT部门提供了一个高效、规范的IT运维管理平台。该系统不仅实现了与目前企业内部使用的业务系统的接口，而且整合了客服、运维和业务管理等系统功能，并可以通过邮件、手机短信等形式对责任人进行阶段提示，提高了系统维护的服务响应效率；通过信息的整合，实现了对各种资源的综合管理，包括各种静态资源、基础资料、备品备件资源的有效管理，从而全面提高了IT部门运行维护的快速响应能力，同时也为IT部门的业务知识积累和业务考核建立了完善的数据模型。

[0137] 语义网：由万维网联盟的蒂姆.伯纳斯-李(Tim Berners-Lee)在1998年提出的一个概念，其实际上是基于很多现有技术的，也依赖于后来和text-and-markup与知识表现的综合。语义网就是能够根据语义进行判断的智能网络，能实现人与电脑之间的无障碍沟通。它好比一个巨型的大脑，智能化程度极高，协调能力非常强大。在语义网上连接的每一部电脑不但能够理解词语和概念，而且还能够理解它们之间的逻辑关系，可以完成人所从事的

工作，能使人类从搜索相关网页的繁重劳动中解放出来。语义网中的计算机能利用自己的智能软件，在万维网上的海量资源中找到所需要的信息，从而将一个个现有的信息孤岛发展成一个巨大的数据库。

[0138] 知识图谱：通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合，并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。它把复杂的知识领域通过数据挖掘、信息处理、知识计量和图形绘制而显示出来，揭示了知识领域的动态发展规律，为学科研究提供切实而有价值的参考。迄今为止，其实际应用在发达国家已经逐步拓展并取得了较好的效果。

[0139] 知识孤岛：由于信息资源得不到有效的交叉融合，知识板块之间相互割裂而形成的无序状态，仿佛大海中的一个个“孤岛”。

[0140] 运维工单：根据不同组织、部门和外部客户的需求，来针对管理，维护和追踪所提出的一系列的问题和请求。一个完善功能的工单系统又可以称为帮助台系统。工单系统一般被广泛用于客户帮助支持服务，客户售后服务，企业IT支持服务，呼叫中心等，用来创建，挂起，解决用户，客户，合作伙伴或企业内部职员提交的事务请求，便于规范化，统一化和清晰化的处理和管理事务。

[0141] (二)构建运维知识图谱的具体实现过程及实现原理

[0142] 本发明专门为运维信息化领域设计了运维知识图谱的生成方法，以解决现有技术无法将知识图谱应用于运维信息化领域的问题。

[0143] 以在运维信息化系统中的应用(运维信息化领域)为例，如图2所示，本发明运维知识图谱的具体构建过程包括：

[0144] (1)信息自动采集：针对运维信息化系统中的运维工单、系统日志等数据源，利用分布式爬虫和接口来提取或抽取出原始的运维知识数据源。

[0145] (2)知识提取/抽取：构建运维知识图谱的第一步，其要解决的关键问题是如何从原始的运维知识数据源这一异构数据源中自动抽取信息来得到候选知识单元。

[0146] 知识提取/抽取过程可进一步细化为：

[0147] Step1：通过一个完整的自然语言分析器抽取可信的三元组$t=(e_i,r_{i,j},e_j)$，并按一定规则将该三元组标记为正值或负值。

[0148] 自然语言分析器首先对运维知识数据源中的所有语句都进行完整的语法分析，形成语法树，并找出每个语句中所有的名词短语e_i，然后通过语法树构建每个语句中所有的名词短语对(e_i,e_j)以及$i<j$间可能存在的相关关系$r_{i,j}$，从而形成一个三元组$t=(e_i,r_{i,j},e_j)$。

[0149] 对每个三元组，可根据这两个名词短语在语法树中是否满足设定的候选判定条件，将其标记为正值或负值。例如，对于一个三元组，若同时满足以下3个条件：1)e_i和e_j之间存在依赖链，且该链长度不超过某个设定数值；2)在语法树中，e_i和e_j并没有跨越句子的界限(例如e_i和e_j并不是一个在主句中出现，而另一个在从句中出现)；3)e_i和e_j都不是代名词(代替某种意义的字或词，例如：“铁公鸡，一毛不拔”，是极端吝啬的代名词)；则这个三元组被标记为正值，反之，则这个三元组被标记为负值。

[0150] Step2：在所有三元组都被标记后，可通过机器学习将这些三元组转换为特征向量

的表示方式，然后将转换后的特征向量作为朴素贝叶斯分类器(用于判断三元组是否可信)的输入，对朴素贝叶斯分类器进行训练。朴素贝叶斯分类器通过计算每一个特征向量正确或错误的频次，最终生成可以被抽取器应用的分类器。

[0151] 具体地，如图3所示，本发明可一次性通过抽取器以三个步骤来实现对需要标注的文档集(即原始的运维知识数据源)的抽取处理：

[0152] 1)利用轻量级的OpenNLP Toolkit对待标注内容中的每条语句进行简单的语法分析，标记出每个词的词性，并识别出名词短语；

[0153] 2)对每对名词短语，如果它们满足设定的候选判定条件，则该对名词短语对应的三元组被标记为候选抽取的三元组；

[0154] 3)利用机器学习方法构造的朴素贝叶斯分类器，对候选抽取的三元组进行分类，如果分朴素贝叶斯类器认为抽取的三元组是可信的，则三元组被抽取出来，存储并归并抽取出来的三元组，使得最终抽取结果中只存储各个不同的三元组和这些三元组出现的频次。

[0155] (3)知识融合：对知识抽取的结果进行清理和整合，消除概念的歧义，剔除冗余和错误的概念，确保知识的质量。

[0156] 如图4所示，知识融合将抽取得到的实体对象链接到知识库中对应的正确实体对象，其具体细化步骤如下：

[0157] Step1：以候选知识单元作为输入文本，对输入文本进行实体识别并生成候选实体。

[0158] Step2：对生成的候选实体进行实体相关性计算并构造相应的实体相关图。其中，顶点构造采用了实体对象，例如：(姚明，姚明(篮球明星))；而边构造则利用了语言相关性计算的结果，语言相关性计算的具体公式如下：

[0159]
$$\begin{cases} w_{ij}=\begin{cases} 1 & e_i\text{和}e_j\text{存在直接关系} \\ GD(e_i,e_j) & e_i\text{和}e_j\text{存在间接关系} \end{cases} \\ GD(e_i,e_j)=1-\dfrac{\log(\max(|in(e_i)|,|in(e_j)|))-\log(|in(e_i)|\cap|in(e_j)|)}{\log(|Kb|)-\log(\min(|in(e_i)|,|in(e_j)|))} \end{cases}$$

[0160] Step3：进行集成化的知识融合。

[0161] 此步骤可进一步细分为：

[0162] 首先，计算候选实体顶点的置信度分数，具体计算公式为：

[0163]
$$PR(v_a)=(1-\alpha)docSim(v_a)+\alpha\sum_{v_b\in Nh(v_a)}\frac{w_{ba}}{\sum_{v_k\in Nh(v_b)}w_{bk}}PR(v_b)$$

[0164] 然后，计算候选实体与输入文本的语义相关性，具体计算公式为：

[0165]
$$SR(v_a,D)=\sum_{v_{k0}\in N_{\max R}}w_{ak0}\cdot PR(v_{k0})$$

[0166] 最后，计算候选实体与实体指称项的语义一致性，具体计算公式为：

[0167] $$SCC(m,c_{k0})=\frac{PR(v_{k0})+SR(v_{k0},D)}{\sum_{V_j\in V_{k0}}PR(v_j)+SR(v_j,D)}$$

[0168] (4)知识加工:对经过知识融合处理的结果进行加工,使其获得结构化、网格化的运维知识图谱体系。

[0169] 如图5所示,知识加工,利用数据驱动的自动化构建方法来进行本体的构建,并进行知识推理,其细化步骤具体包括:

[0170] Step1:对知识融合结果进行实体并列关系相似度计算。运维知识实体间的并列关系相似度为考察任意给定的2个运维知识实体在多大程度上属于同一概念分类的指标测度,相似度越高,表明这2个运维知识实体越有可能属于同一语义类别。例如“中国”和“美国”作为国家名称的实体,具有较高的并列关系相似度,属于同一语义类别的可能性较高;而“中国”和“苹果”这两个实体,具有较低的并列关系相似度,属于同一语义类别的可能性较低。具体计算实体并列关系相似度时,首先将每个运维知识实体表示成1个N维向量(其中,向量的每个维度表示1个预先定义的上下文环境,向量元素值表示该运维知识实体出现在各上下文环境中的概率),然后就可以通过求解向量间的相似度来得到运维知识实体间的并列关系相似度。

[0171] Step2:进行实体上下级关系抽取,以确定概念之间的隶属(IsA)关系,如确定词组(汽车,交通工具)构成的上下级关系。

[0172] Step3:本体的生成。运维知识实体的本体的生成具体过程为:对各层次得到的概念(即实体上下级关系)进行聚类,并对其进行语义类的标定(如为该类中的实体指定1个或多个公共上级词等)。

[0173] Step4:知识推理。知识推理的具体过程为:从已有的实体关系数据出发,经过计算机推理,建立运维知识实体间的新关联,从而得到对应的运维知识图谱。

[0174] 例如已知(张三,上级,李四)和(李四,上级,王五),可以通过知识推理得到(张三,上级,王五)或(王五,下级,张三)。此处知识推理算法的基本思想是将运维知识图谱视为图(以运维知识实体为节点,以关系或属性为边),从源节点开始,在图上执行随机游走操作,若能够通过一条路径到达目标节点,则推测源节点和目的节点间可能存在关系。例如:假设2个节点(X,Y)共有1个孩子Z,即存在路径X→Z←Y,据此可推测X和Y之间可能存在“MarriedTo(婚姻)”关系。

[0175] (5)知识更新:随着时间的演进,不断对运维知识图谱进行迭代更新,保持运维知识库的与时俱进。

[0176] 知识更新的具体细化步骤为:

[0177] Step1:通过图2的信息自动采集过程实时获取新的运维知识数据源。

[0178] Step2:对新的运维知识数据源进行预处理,并将预处理后的运维知识数据源中的数据分别标记为第一数据A和第二数据B。其中,预处理,用于对新的运维知识数据源中的数据进行规则检查和过滤,去除冗余的信息。第一数据A是指与现有运维知识图谱的数据的差异大于设定的差异阈值的数据,即完全不同的数据。第二数据B是指与现有运维知识图谱的数据的差异小于等于设定的差异阈值的数据,即有较小差异的数据。

[0179] Step3:以第一数据A作为增量数据,然后依次执行图2的知识抽取、知识融合和知

识加工操作，最终得到第一数据对应的新运维知识图谱，并将第一数据A对应的新运维知识图谱补充到用于存储运维知识图谱数据的运维知识图谱数据库中，以丰富运维知识图谱。

[0180] Step4：分析出第二数据B中区别于现有运维知识图谱的运维知识实体和第一运维知识实体关系C(即分析出第二数据B与现有知识图谱相矛盾的运维知识实体和关系C)，并判断第二数据B的时序性是否小于1，若是，则将第二数据B从预处理后的运维知识数据源中剔除，反之，则将第一运维知识实体关系C标记为历史数据，然后对第一运维知识实体关系C的时序性进行计算和排序，并根据计算和排序的结果更新现有运维知识图谱。本发明在进行时序性判断和排序时，会标记过期的实体或关系为历史数据，从而在不影响当前真实运维知识图谱的展示下提供溯源服务来对运维知识进行溯源，更加方便。

[0181] (6)知识存储：对运维知识及运维知识间的相互关系信息进行存储。

[0182] 本实施例实现了在运维信息化系统中基于运维工单等信息来构建运维知识图谱的目的，并能在后续运行过程中不断获取新的运维数据来对运维知识图谱进行丰富和演进。

[0183] 与现有技术相比，本发明具有以下优点：

[0184] (a)基于语义分析和图论的模型构建了应用于运维信息化领域的运维知识图谱，整个运维知识图谱的创建过程除了最初的规则制定和图谱生成后的人为审核外，无需其它人力投入，成本更低，效率更高。

[0185] (b)实现了运维知识图谱的自适应：通过实时获取新的运维知识数据，自动对运维知识图谱中已有的数据进行增量更新和修正(包括删除实体，取消关系等操作)，实时性高且更加方便；

[0186] (c)实现了运维知识图谱的演化：能够自动根据已知的知识关系通过知识推理推演出新的知识关系，并回馈到运维知识图谱中丰富其构成，更加方便；同时，运维知识图谱中旧的关系也不会彻底删除，而是作为历史数据(即知识“历史”)进行存储，以便后续进行知识的溯源操作。

[0187] (d)关联准确率高：基于运维知识图谱的自适应和演化特性，能够不断对运维知识图谱进行修正，准确性更高。

[0188] 以上是对本发明的较佳实施进行了具体说明，但本发明并不限于所述实施例，熟悉本领域的技术人员在不违背本发明精神的前提下还可做作出种种的等同变形或替换，这些等同的变形或替换均包含在本申请权利要求所限定的范围内。

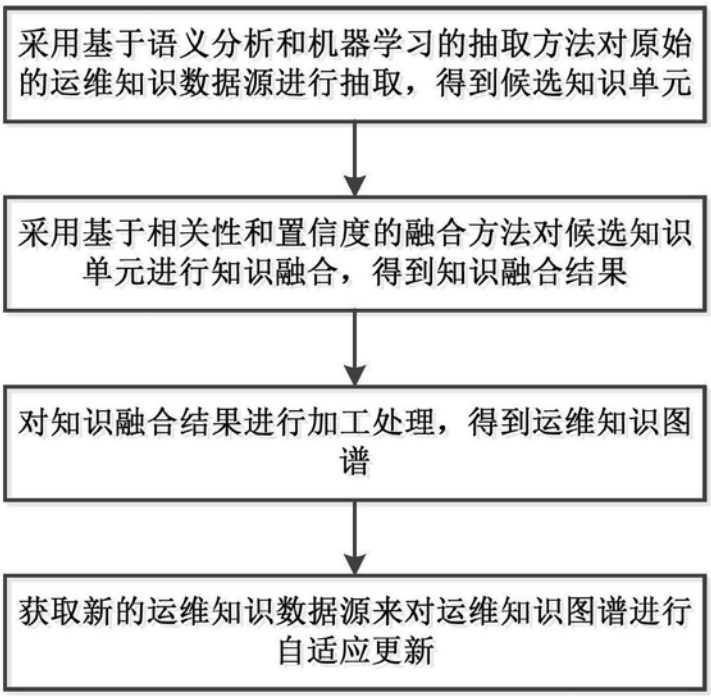

采用基于语义分析和机器学习的抽取方法对原始的运维知识数据源进行抽取，得到候选知识单元
采用基于相关性和置信度的融合方法对候选知识单元进行知识融合，得到知识融合结果
对知识融合结果进行加工处理，得到运维知识图谱
获取新的运维知识数据源来对运维知识图谱进行自适应更新

图1

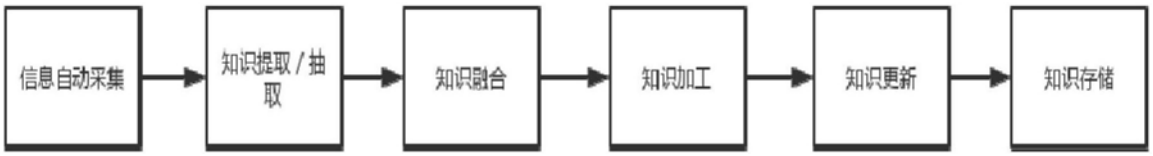

信息自动采集
知识提取/抽取
知识融合
知识加工
知识更新
知识存储

图2

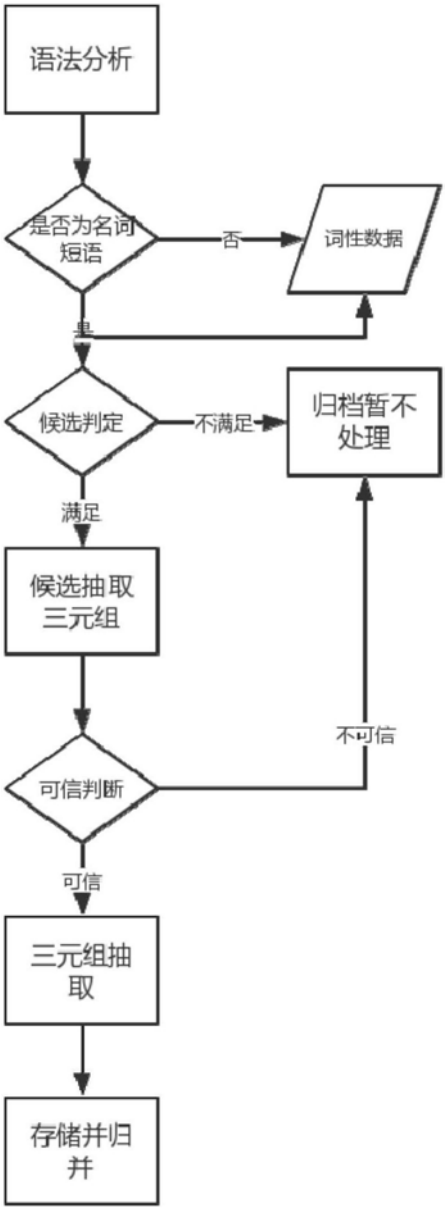

语法分析
是否为名词短语
否
词性数据
是
候选判定
不满足
归档暂不处理
满足
候选抽取三元组
可信判断
不可信
可信
三元组抽取
存储并归并

图3

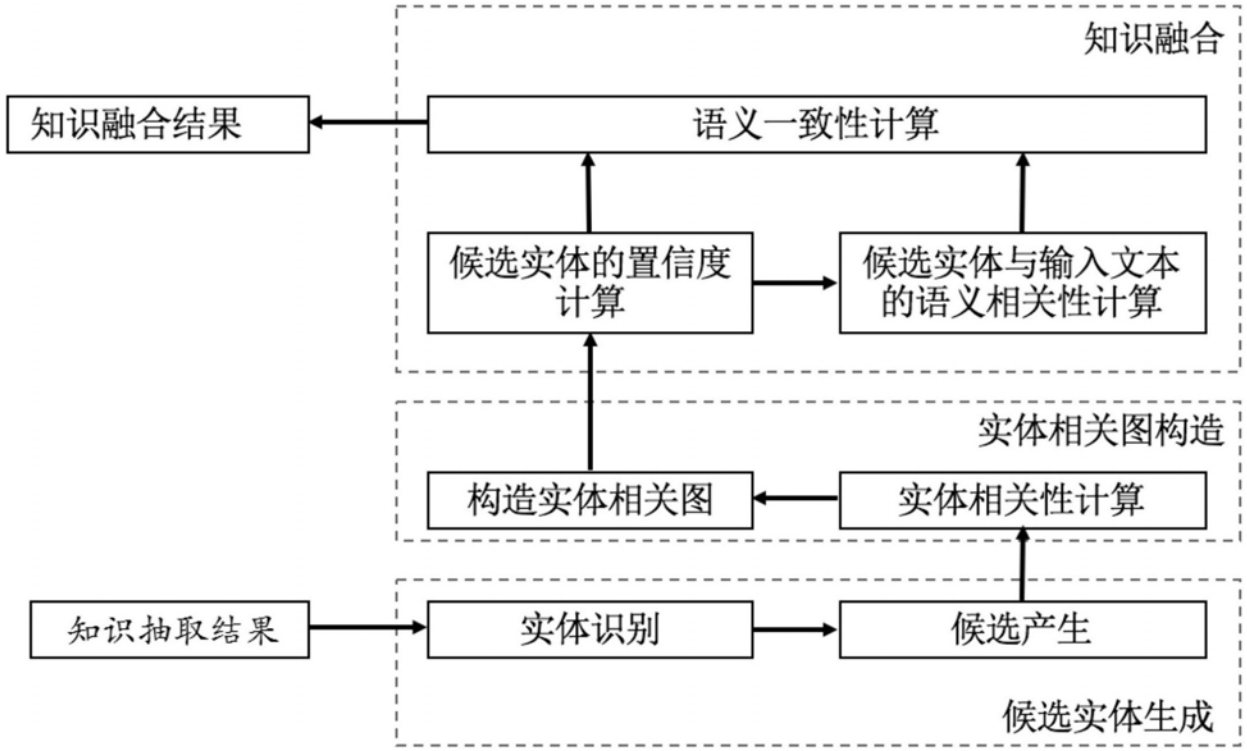

知识融合
知识融合结果
语义一致性计算
候选实体的置信度计算
候选实体与输入文本的语义相关性计算
实体相关图构造
构造实体相关图
实体相关性计算
知识抽取结果
实体识别
候选产生
候选实体生成

图4

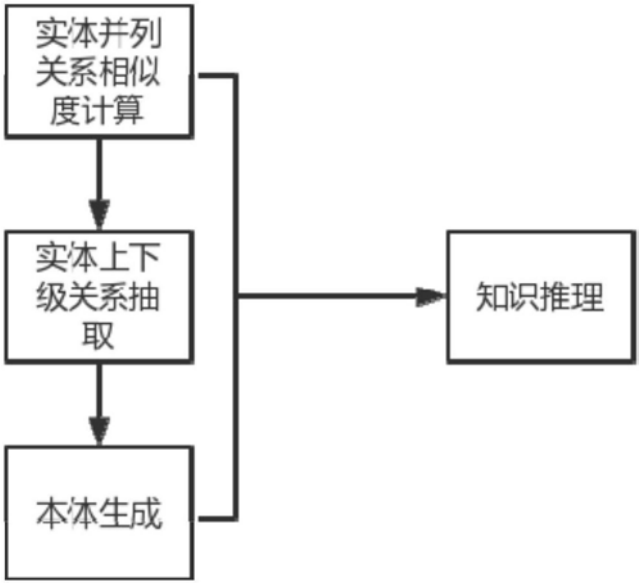

实体并列关系相似度计算
实体上下级关系抽取
本体生成
知识推理

图5