

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 7 月 16 日 (16.07.2020)



(10) 国际公布号
WO 2020/143184 A1

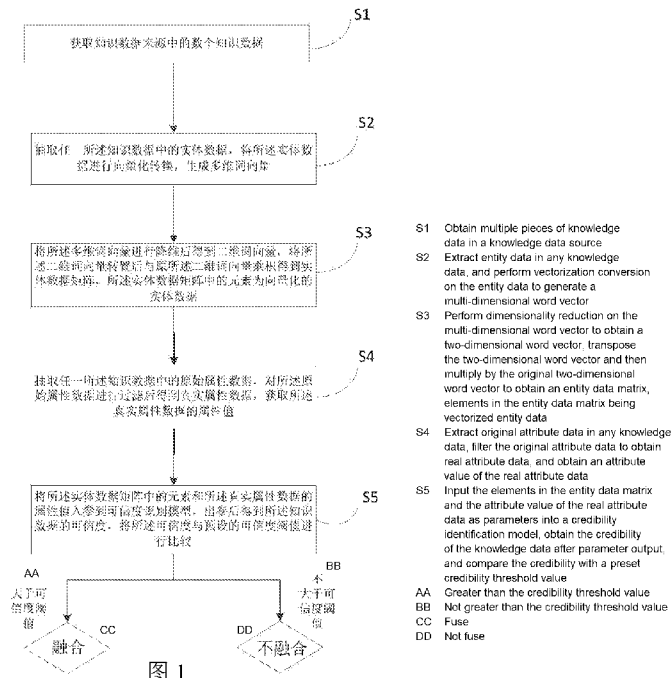
- (51) 国际专利分类号:
G06K 9/62 (2006.01) **G06F 16/36** (2019.01)
G06F 16/951 (2019.01)
- (21) 国际申请号: PCT/CN2019/092597
- (22) 国际申请日: 2019 年 6 月 24 日 (24.06.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910025114.4 2019 年 1 月 11 日 (11.01.2019) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

- (72) 发明人: 孙佳兴(SUN, Jiaying); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。 胡逸凡(HU, Yifan); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。 陈泽晖(CHEN, Zehui); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。 黄鸿顺(NG, Jon Tzen); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(54) Title: KNOWLEDGE FUSION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 知识融合方法、装置、计算机设备和存储介质



(57) Abstract: The present application relates to the technical field of knowledge graphs, in particular, to a knowledge fusion method and apparatus, a computer device, and a storage medium. The method comprises: obtaining multiple pieces of knowledge data in a knowledge data source; extracting entity data in any knowledge data, and performing vectorization conversion on the entity data to generate a multi-dimensional word vector; performing dimensionality reduction on the multi-dimensional word vector to obtain a two-dimensional word vector, transposing the two-dimensional word vector and then multiplying by the original two-dimensional word

(74) 代理人:北京市京大律师事务所(BEIJING JINGDA LAW FIRM); 中国北京市海淀区海淀中街16号7层2单元703, Beijing 100080 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

vector to obtain an entity data matrix, elements in the entity data matrix being vectorized entity data; obtaining an attribute value of real attribute data; and inputting the elements in the entity data matrix and the attribute value of the real attribute data as parameters into a credibility identification model, obtaining the credibility of the knowledge data after parameter output, comparing the credibility with a preset credibility threshold value, and performing fusion. The present application achieves effective fusion of multiple attributes in the same entity.

(57) 摘要: 本申请涉及知识图谱技术领域, 尤其涉及一种知识融合方法、装置、计算机设备和存储介质, 包括: 获取知识数据来源中的数个知识数据; 抽取任一所述知识数据中的实体数据, 将所述实体数据进行向量化转换, 生成多维词向量; 将所述多维词向量进行降维后得到二维词向量, 将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵, 所述实体数据矩阵中的元素为向量化的实体数据; 获取所述真实属性数据的属性值; 将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型, 出参后得到所述知识数据的可信度, 将所述可信度与预设的可信度阈值进行比较后进行融合。本申请实现了同一实体中多个属性的有效融合。

知识融合方法、装置、计算机设备和存储介质

本申请要求于2019年1月11日提交中国专利局、申请号为201910025114.4、发明名称为“知识融合方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及知识图谱技术领域，尤其涉及一种知识融合方法、装置、计算机设备和存储介质。

背景技术

在当今互联网上存在着大量的知识，在各个网页包含的数据信息中存在各种样式的知识数据。其中，知识数据有三个部分组成，分别是：实体信息、关系信息和属性信息。在对知识数据进行梳理时，需要对知识数据进行融合，这一过程称为知识融合。

知识融合，是指发现异构数据库中相同概念的不同表达，它通过对分布式数据源和知识源进行组织和管理，结合应用需求对知识元素进行转化、集成和融合等处理，从而获取有价值或可用的新知识，同时对知识对象的结构和内涵进行优化，提供基于知识的服务。知识融合的研究对于分布式知识库环境中的知识共享、知识系统的交互、集成和协同工作、知识服务质量的优化等方面具有一定的价值，对于研究基于知识内涵的知识发现以及新知识的创建、组织、评价和优化等方面具有相当重要的意义。

目前，在进行知识融合的过程中存在着不能对属性进行精准判断，从而导致在融合的过程中，无法将属于同一实体的多个属性的有效进行合并的问题。

发明内容

基于此，有必要针对无法将属于同一实体的多个属性的有效进行合并的问题，提供一种知识融合方法、装置、计算机设备和存储介质。

25 一种知识融合方法，包括：

获取知识数据来源中的数个知识数据；抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；将所述实体数据矩阵中的元素和

所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

一种知识融合装置，包括如下模块：

5 数据获取模块，设置为获取知识数据来源中的数个知识数据；

向量生成模块，设置为抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；

数据向量化模块，设置为将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；

属性值获取模块，设置为抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；

15 融合判定模块，设置为将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述知识融合方法的步骤。

一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行上述知识融合方法的步骤。

上述知识融合方法、装置、计算机设备和存储介质，包括：获取知识数据来源中的数个知识数据；抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。本技术方案通过对实体和属性的精确匹配，

实现了同一实体的多个属性的有效融合。

附图说明

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，而并不认为是对本申请的限制。

图 1 为本申请在一个实施例中的一种知识融合方法的整体流程图；

图 2 为本申请在一个实施例中的一种知识融合方法的数据获取过程示意图；

图 3 为本申请在一个实施例中的一种知识融合方法的向量生成过程示意图；

图 4 为本申请在一个实施例中的一种知识融合装置的结构图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本申请的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。

图 1 为本申请在一个实施例中的一种知识融合方法的整体流程图，如图 1 所示，一种知识融合方法，包括：

S1，获取知识数据来源中的数个知识数据；

具体的，本步骤中的知识数据可以来自于同一知识数据来源，也可以来自于不同数据来源，可以来自于本地数据，也可以来自网络数据。若来自于本地数据，则在获取知识数据时，需要获得知识数据的存储路径；若来自于网络数据来源，则在获取知识数据时，需要获得知识数据来源的网络地址。

S2，抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；

具体的，获取存储在数据库中的实体名称列表，随机抽取实体名称列表中的至少一个实体名称，根据所述实体名称从所述知识数据中抽取出实体数据。同时，在进行实体数据抽取时，可以采用近义词抽取的方法；比如在实体名称列表中抽取出的实体名称是“篮球”，那么在对所述知识数据中的实体数据进行抽取时，可以抽取出的实体数据的实体名称为“足球”、“排球”等球类运动名词。

S3，将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；

具体的，在对多维词向量进行降维时可以采用 PCA 降维，比如，设有 m 条 n 维数据，降维操作可以采用如下步骤：

1) 将原始数据按列组成 n 行 m 列形成矩阵 X ；2) 将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值；3) 求出 X 矩阵的协方差矩阵 Y ；4) 求出协方差矩阵 Y 的特征值及对应的特征向量 r ；5) 将特征向量按对应特征值大小从上到下按行排列成矩阵 Z ，取前 k 行组成矩阵 Q ；6) 矩阵 Q 即为降维到 k 维后的数据。

S4，抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；

具体的，在进行过滤时，主要是过滤掉与语义无关的词语。可以将原始属性数据进行分割，分割成数个子数据段，然后对每一个子数据段中的数据进行属性词查询，若不存在属性词，则将所述子数据段清除。

S5，将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

具体的，可信度即可靠性，它指的是采取同样的方法对同一对象重复进行测量时，其所得结果相一致的程度。从另一方面来说，可信度就是指测量数据的可靠程度。其中，预设的可信度阈值是根据历史数据统计后得到的，一般可信度阈值设定为 95%。

本实施例，通过对实体数据和属性数据进行有效加工，从而实现了同一实体的多个属性的有效融合。

图2为本申请在一个实施例中的一种知识融合方法的数据获取过程示意图，如图所示，所述S1，获取知识数据来源中的数个知识数据，包括：

S101、发送知识数据抽取指令至待抽取知识数据的所述知识数据来源；

具体的，获取所述待抽取知识数据的知识数据来源的网络地址，根据所述网络地址的格式确定所述网络地址的类型，即所述网络地址是静态IP地址还是动态IP地址，若是静态IP地址，则从数据库中调取IP地址表进行比对后，确定所述静态IP地址是否在所述IP地址表上，在则发送知识数据获取指令，不在则不发送；若是动态IP地址，则对所述动态IP地址进行DNS解析得到所述动态IP地址对应的DNS解析代码，而后调用数据库中的DNS解析代码表对所述DNS解析代码进行比对，确定所述DNS解析代码是否在所述DNS解析代码表上，在则发送知识数据获取指令，不在则不发送。

S102、接收所述知识数据来源的反馈信息，从所述反馈信息中抽取出数据来源类型的关键词，根据所述关键词确定所述知识来数据源的类型；

具体的，形式关键词是指知识数据是结构化数据、半结构化数据还是非结构化数据。比如，反馈信息中出现“表”这一形式关键词，则对应的是结构化数据；出现“网页”这一形式关键词，则对应的是半结构化数据；出现“文本”这一形式关键词，则对应非结构化数据。

S103、获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据。

具体的，不同形式的数据来源对应不同的数据抽取方法，比如，半结构化的网页信息通常采用网络爬虫进行爬取，对于非结构化的文本，通常采用文本语言进行抽取。

本实施例，通过对知识数据来源的反馈信息进行分析，确定知识数据来源的数据形式，从而能够采用正确的抽取方式对知识数据来源的知识数据进行抽取。

图3为本申请在一个实施例中的一种知识融合方法的向量生成过程示意图，如图所示，所述S2，抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量，包括：

S201、设置一抽取所述知识数据中的实体数据的初始语段，所述初始语段中至少包含一个所述实体数据；

具体的，根据实体数据中实体词语的长度值历史数据，设定实体数据的初始语段的长度。比如，在数据库中存储的历史数据中，实体词语的长度从 1~10，那么初始语段的长度设置为最大值 10。

5 S202、根据所述初始语段的语段长度，将所述知识数据分割成数个初始子数据块，若任意一个所述初始子数据块中包含有两个或者以上的实体数据，则将所述初始子数据块进行再次分割得到只包含一个所述实体数据的最终子数据块；

具体的，在对初始语段进行分割时，每一个子数据块的长度可以不一致，即根据实际实体词语的长度确定每一个子数据块的长度。

10 S203、抽取所述最终子数据块中的实体数据，提取所述最终子数据块中的实体数据的语义特征，应用词向量转化法将所述语义特征转换成初始多维词向量，将所述最终子数据块的语段长度作为系数与所述初始多维词向量进行乘积后得到最终多维词向量。

具体的，语义特征包含语义、语法、结构等多方面的要素。词向量转化法通常使用的是 Word2Vector 算法，此算法可以对每一个语义特征进行上下联系，从而将相互关联的语义特征一起转换成初始多维词向量。

本实施例，通过实体数据进行向量化转换后，使实体数据数值化表示，便于使用机器学习方法进行相似性计算。

20 在一个实施例中，所述 S3，将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据，包括：

获取所述多维词向量中的每一个样本点的 K 个近邻点；

具体的，样本点是指多维向量中的每一个点；在多维空间 N 上每一个样本点存在着在同一平面上的直接连接的点，这些点成为近邻点，K 的取值范围为 25 1~n，n 为非零正整数。

根据每个样本点的 K 个近邻点，建立所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ；

根据所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ，将每个样本点映射到低维空间，映射条件为：

30
$$\min \varepsilon(Y) = \sum_{i=1}^N \left\| y_n - \sum_{j=1}^K w_{ij} y_{ij} \right\|$$
，式中： $\varepsilon(Y)$ 为损失函数值， y_{ij} 为近邻点数值， y_n

为近邻点输出向量， W_{ij} 是局部权重矩阵中的元素，K 为近邻点个数，N 为近邻

点输出向量中元素的个数，映射后得到二维词向量 $Y=\{y_1, y_2, \dots, y_N\}$ ；将所述二维词向量转置后得到转置二维词向量，将所述二维词向量与所述转置二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据。

5 本实施例，通过将多维词向量进行降维成二维词向量，从而方便实体信息和属性信息进行匹配。

在一个实施例中，所述 S4、抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值，包括：

10 抽取所述任一所述知识数据中的原始属性数据，离散化处理所述原始属性数据后得到所述原始属性数据的离散值；

具体的，离散化是指把无限空间中有限的个体映射到有限的空间中去，以此来提高算法的时空效率。在进行离散化处理前，可以使用如 `unique()` 去重函数去除知识数据中的重复数据，而后再对知识数据离散化，其中，`unique()` 函数是一种被 C++、PHP、Matlab 等开发或者科学计算环境所支持的去重函数，用于
15 去除集合中的重复值，或者从集合中取单值。

根据所述知识数据中的所述原始属性数据的数量，获取所述原始属性数据对应的向量维度；

其中，原始属性数据的向量维度等于原始属性数据的数量。

20 将所述离散值和所述向量维度进行做差，若差值在预设的误差阈值以内，则所述原始属性数据为真实属性数据，若所述差值不在所述误差阈值以内，则根据所述差值去除所述原始属性数据中的多余属性数据，得到所述真实属性数据；

根据所述真实属性数据的数量，获取所述真实属性数据对应的向量维度，建立真实属性数据向量；

25 将所述真实属性数据向量降维后形成一真实属性数据矩阵，获取所述真实属性数据矩阵的特征值，所述特征值为所述属性值。

具体的，将所述真实属性数据降维后，可以得到一二维属性向量，将二维属性向量转置后得到转置二维属性向量，将二维属性向量和转置二维属性向量乘积后得到真实属性向量。

本实施例，通过对原始属性数据进行降维并进行矩阵化处理，更好的得到真实的属性值。

在一个实施例中，所述 S5，将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合，包括：

获取任一所述实体数据矩阵中的元素和任一所述真实属性数据的属性值，将所述元素和所述属性值入参到相似度距离函数中计算相似度距离，计算公式：

$$L(m_1, m_2) = \sqrt{\frac{1}{2} (\|\vec{m_1}\|^2 + \|\vec{m_2}\|^2 - 2\langle \vec{m_1}, \vec{m_2} \rangle)}$$
 式中：L(m₁, m₂)为相似度距离函

数，m₁为元素，m₂为属性值；

根据所述相似度距离，计算得到所述元素和所述属性值的可信度，计算公式为：

$$Crd(m) = \frac{L(m_1, m_2)}{\sum_{i=1}^N 1 - L(m_1, m_2)}$$
 式中：Crd(m)为可信度函数，L(m₁, m₂)为相似度距离函

数；将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将

抽取出的同一实体数据对应的所述原始属性数据融合，否则不融合。

具体的，在进行相似度计算时还可以采用余弦算法或者欧式距离算法等，可信度阈值是根据历史数据统计后得到的。

本实施例，通过对实体数据和属性数据可信度的计算，从而提升了属性数据融合的准确性。

在一个实施例中，所述 S103、获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据包括：

若所述获取所述知识数据来源的形式为网页，则应用网络爬虫工具进行抽取，包括：

获取预抽取知识数据的任务队列中的关键词组，所述关键词组中包含有多个关键词；其中，任务队列中的关键词组可以是一些性状性的词组，比如：“球类”，在这个关键词组下所包含的关键词可以有“篮球”、“足球”、“乒乓球”等。

遍历所述关键词组，通过网络爬虫爬取与所述关键词组中每一个关键词对应的网页上的信息；获取所述网页上的信息中的所有实体信息，将所述实体信息导入到预设的知识数据表中，若有一个及以上的实体信息无法导入到预设的

知识数据表中，则重新通过网络爬虫爬取网页，否则将所述网页信息作为所述知识数据。

具体的，实体信息是指实体名称等和“实体”相关的信息，在导入到预设的知识数据表中时，先检索所述预设的知识数据表中的实体名称，若某一个实体信息中的实体名称不在预设的知识数据表中，则无法将所述实体信息导入。其中，

预设的知识数据表存储在数据库中，其根据历次知识数据采集后汇总而成。

本实施例，能够有效的从网页信息中抽取所需要的知识数据，提升知识

数据抽取的效率。

在一个实施例中，提出了一种知识融合装置，如图4所示，包括如下模块：

数据获取模块41，设置为获取知识数据来源中的数个知识数据；

向量生成模块42，设置为抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；

数据向量化模块43，设置为将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；

属性值获取模块44，设置为抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；

融合判定模块45，设置为将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

在一个实施例中，所述向量生成模块，还设置为：

发送知识数据抽取指令至待抽取知识数据的所述知识数据来源；接收所述知识数据来源的反馈信息，从所述反馈信息中抽取数据来源类型的关键词，根据所述关键词确定所述知识数据来源的类型；获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据。

在一个实施例中，所述数据获取模块，还设置为：

设置一抽取所述知识数据中的实体数据的初始语段，所述初始语段中至少包含一个所述实体数据；根据所述初始语段的语段长度，将所述知识数据分割成数个初始子数据块，若任意一个所述初始子数据块中包含有两个或者以上的实体数据，则将所述初始子数据块进行再次分割得到只包含一个所述实体数据的最终子数据块；抽取所述最终子数据块中的实体数据，提取所述最终子数据

块中的实体数据的语义特征，应用词向量转化法将所述语义特征转换成初始多维词向量，将所述最终子数据块的语段长度作为系数与所述初始多维词向量进行乘积后得到最终多维词向量。

在一个实施例中，所述数据向量化模块，还设置为：

- 5 获取所述多维词向量中的每一个样本点的 K 个近邻点；根据每个样本点的 K 个近邻点，建立所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ；

根据所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ，将每个样本点映射到低维空间，映射后得到二维词向量 $Y = \{y_1, y_2, \dots, y_N\}$ ；将所述二维词向量转置后得到转置二维词向量，将所述二维词向量与所述转置二维词向量乘积得到
10 实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据。

在一个实施例中，所述属性值获取模块，还设置为设置为：

抽取所述任一所述知识数据中的原始属性数据，离散化处理所述原始属性数据后得到所述原始属性数据的离散值；根据所述知识数据中的所述原始属性数据的数量，获取所述原始属性数据对应的向量维度；将所述离散值和所述向
15 量维度进行做差，若差值在预设的误差阈值以内，则所述原始属性数据为真实属性数据，若所述差值不在所述误差阈值以内，则根据所述差值去除所述原始属性数据中的多余属性数据，得到所述真实属性数据；根据所述真实属性数据的数量，获取所述真实属性数据对应的向量维度，建立真实属性数据向量；

将所述真实属性数据向量降维后形成一真实属性数据矩阵，获取所述真实
20 属性数据矩阵的特征值，所述特征值为所述属性值。

在一个实施例中，所述融合判定模块，还设置为：

获取任一所述实体数据矩阵中的元素和任一所述真实属性数据的属性值，将所述元素和所述属性值入参到相似度距离函数中计算相似度距离；根据所述相似度距离，计算得到所述元素和所述属性值的可信度；将所述可信度与预设
25 的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的同一实体数据对应的所述原始属性数据融合，否则不融合。

在一个实施例中，所述向量生成模块，还设置为：

获取预抽取知识数据的任务队列中的关键词组，所述关键词组中包含有多个关键词；遍历所述关键词组，通过网络爬虫爬取与所述关键词组中每一个关
30 键词对应的网页上的信息；获取所述网页上的信息中的所有实体信息，将所述实体信息导入到预设的知识数据表中，若有一个及以上的实体信息无法导入到预设的知识数据表中，则重新通过网络爬虫爬取网页，否则将所述网页信息作为所述知识数据。

在一个实施例中，提出了一种计算机设备，所述计算机设备包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行上述各实施例中所述知识融合方法的步骤。

5 在一个实施例中，提出了一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行上述各实施例中所述知识融合方法的步骤。所述存储介质可以为非易失性存储介质。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：只读存储器(ROM, Read Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁盘或光盘等。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

15 以上所述实施例仅表达了本申请一些示例性实施例，其中描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权 利 要 求 书

1、一种知识融合方法，其中，包括：

获取知识数据来源中的数个知识数据；抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

2、根据权利要求 1 所述的知识融合方法，其中，所述获取知识数据来源中的数个知识数据，包括：

发送知识数据抽取指令至待抽取知识数据的所述知识数据来源；接收所述知识数据来源的反馈信息，从所述反馈信息中抽取数据源类型的关键词，根据所述关键词确定所述知识数据来源的类型；获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据。

3、根据权利要求 1 所述的知识融合方法，其中，所述抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量，包括：

设置一抽取所述知识数据中的实体数据的初始语段，所述初始语段中至少包含一个所述实体数据；根据所述初始语段的语段长度，将所述知识数据分割成数个初始子数据块，若任意一个所述初始子数据块中包含有两个或者以上的实体数据，则将所述初始子数据块进行再次分割得到只包含一个所述实体数据的最终子数据块；抽取所述最终子数据块中的实体数据，提取所述最终子数据块中的实体数据的语义特征，应用词向量转化法将所述语义特征转换成初始多维词向量，将所述最终子数据块的语段长度作为系数与所述初始多维词向量进行乘积后得到最终多维词向量。

4、根据权利要求 1 所述的知识融合方法，其中，所述将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据，包括：

获取所述多维词向量中的每一个样本点的 K 个近邻点；根据每个样本点的 K 个近邻点，建立所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ；根据所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ，将每个样本点映射到低维空间，映射条件为：

$$\min \varepsilon(Y) = \sum_{i=1}^N \left\| y_n - \sum_{j=1}^K w_{ij} y_{ij} \right\|, \text{ 式中: } \varepsilon(Y) \text{ 为损失函数值, } y_{ij} \text{ 为近邻点数值, } y_n \text{ 为}$$

近邻点输出向量, w_{ij} 是局部权重矩阵中的元素, K 为近邻点的个数, N 为近邻点输出向量中元素的个数, 映射后得到二维词向量 $Y=\{y_1, y_2, \dots, y_N\}$; 将所述二维词向量转置后得到转置二维词向量, 将所述二维词向量与所述转置二维词向量

5 乘积得到实体数据矩阵, 所述实体数据矩阵中的元素为向量化的实体数据。

5、根据权利要求 1 所述的知识融合方法, 其中, 所述抽取任一所述知识数据中的原始属性数据, 对所述原始属性数据进行过滤后得到真实属性数据, 获取所述真实属性数据的属性值, 包括:

抽取所述任一所述知识数据中的原始属性数据, 离散化处理所述原始属性
10 数据后得到所述原始属性数据的离散值; 根据所述知识数据中的所述原始属性数据的数量, 获取所述原始属性数据对应的向量维度; 将所述离散值和所述向量维度进行做差, 若差值在预设的误差阈值以内, 则所述原始属性数据为真实属性数据, 若所述差值不在所述误差阈值以内, 则根据所述差值去除所述原始属性数据中的多余属性数据, 得到所述真实属性数据; 根据所述真实属性数据
15 的数量, 获取所述真实属性数据对应的向量维度, 建立真实属性数据向量; 将所述真实属性数据向量降维后形成一真实属性数据矩阵, 获取所述真实属性数据矩阵的特征值, 所述特征值为所述属性值。

6、根据权利要求 1 所述的知识融合方法, 其中, 所述将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型, 出参后得到所
20 述知识数据的可信度, 将所述可信度与预设的可信度阈值进行比较, 若大于所述可信度阈值则将抽取出的所述原始属性数据融合, 否则不融合, 包括:

获取任一所述实体数据矩阵中的元素和任一所述真实属性数据的属性值, 将所述元素和所述属性值入参到相似度距离函数中计算相似度距离, 计算公式
为: $L(m_1, m_2) = \sqrt{\frac{1}{2} \left(\left\| \vec{m_1} \right\|^2 + \left\| \vec{m_2} \right\|^2 - 2 \langle \vec{m_1}, \vec{m_2} \rangle \right)}$, 式中: $L(m_1, m_2)$ 为相似度距离函数, m_1 为
25 元素, m_2 为属性值; 根据所述相似度距离, 计算得到所述元素和所述属性值的可信度, 计算公式为:

$$\text{Crd}(m) = \frac{L(m_1, m_2)}{\sum_{i=1}^N 1 - L(m_1, m_2)}, \text{ 式中: } \text{Crd}(m) \text{ 为可信度函数, } L(m_1, m_2) \text{ 为相似度距离}$$

函数；将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的同一实体数据对应的所述原始属性数据融合，否则不融合。

7、根据权利要求 2 所述的知识融合方法，其中，所述获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据包括：若所述获取所述知识数据来源的形式为网页，则应用网络爬虫工具进行抽取，包括：获取预抽取知识数据的任务队列中的关键词组，所述关键词组中包含有多个关键词；遍历所述关键词组，通过网络爬虫爬取与所述关键词组中每一个关键词对应的网页上的信息；获取所述网页上的信息中的所有实体信息，将所述实体信息导入到预设的知识数据表中，若有一个及以上的实体信息无法导入到预设的知识数据表中，则重新通过网络爬虫爬取网页，否则将所述网页信息作为所述知识数据。

8、一种知识融合装置，其中，包括：

数据获取模块，设置为获取知识数据来源中的数个知识数据；

15 向量生成模块，设置为抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；

数据向量化模块，设置为将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；

20 属性值获取模块，设置为抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；

融合判定模块，设置为将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

9、根据权利要求 8 所述的知识融合装置，其中，所述向量生成模块，还设置为：

30 发送知识数据抽取指令至待抽取知识数据的所述知识数据来源；接收所述知识数据来源的反馈信息，从所述反馈信息中抽取数据源类型的关键词，

根据所述关键词确定所述知识来数据源的类型；获取所述知识数据来源的类型对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据。

10、根据权利要求 8 所述的知识融合装置，其中，所述数据获取模块，还设置为：

5 设置一抽取所述知识数据中的实体数据的初始语段，所述初始语段中至少包含一个所述实体数据；根据所述初始语段的语段长度，将所述知识数据分割成数个初始子数据块，若任意一个所述初始子数据块中包含有两个或者以上的实体数据，则将所述初始子数据块进行再次分割得到只包含一个所述实体数据的最终子数据块；抽取所述最终子数据块中的实体数据，提取所述最终子数据块中的实体数据的语义特征，应用词向量转化法将所述语义特征转换成初始多维词向量，将所述最终子数据块的语段长度作为系数与所述初始多维词向量进行乘积后得到最终多维词向量。

11、根据权利要求 8 所述的知识融合装置，其中，所述数据向量化模块，还设置为：

15 获取所述多维词向量中的每一个样本点的 K 个近邻点；根据每个样本点的 K 个近邻点，建立所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ；

 根据所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ，将每个样本点映射到低维空间，映射后得到二维词向量 $Y = \{y_1, y_2, \dots, y_N\}$ ；将所述二维词向量转置后得到转置二维词向量，将所述二维词向量与所述转置二维词向量乘积得到
20 实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据。

12、根据权利要求 8 所述的知识融合装置，其中，所述属性值获取模块，还设置为：

 抽取所述任一所述知识数据中的原始属性数据，离散化处理所述原始属性数据后得到所述原始属性数据的离散值；根据所述知识数据中的所述原始属性
25 数据的数量，获取所述原始属性数据对应的向量维度；将所述离散值和所述向量维度进行做差，若差值在预设的误差阈值以内，则所述原始属性数据为真实属性数据，若所述差值不在所述误差阈值以内，则根据所述差值去除所述原始属性数据中的多余属性数据，得到所述真实属性数据；根据所述真实属性数据的数量，获取所述真实属性数据对应的向量维度，建立真实属性数据向量；
30 将所述真实属性数据向量降维后形成一真实属性数据矩阵，获取所述真实属性数据矩阵的特征值，所述特征值为所述属性值。

13、根据权利要求 8 所述的知识融合装置，其中，所述融合判定模块，还设置为：

获取任一所述实体数据矩阵中的元素和任一所述真实属性数据的属性值，将所述元素和所述属性值入参到相似度距离函数中计算相似度距离；根据所述相似度距离，计算得到所述元素和所述属性值的可信度；将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的同一实体数据对应的所述原始属性数据融合，否则不融合。

14、根据权利要求9所述的知识融合装置，其中，所述向量生成模块，还设置为：

获取预抽取知识数据的任务队列中的关键词组，所述关键词组中包含有多个关键词；遍历所述关键词组，通过网络爬虫爬取与所述关键词组中每一个关键词对应的网页上的信息；获取所述网页上的信息中的所有实体信息，将所述实体信息导入到预设的知识数据表中，若有一个及以上的实体信息无法导入到预设的知识数据表中，则重新通过网络爬虫爬取网页，否则将所述网页信息作为所述知识数据。

15、一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取知识数据来源中的数个知识数据；抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实属性数据，获取所述真实属性数据的属性值；将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

16、一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

获取知识数据来源中的数个知识数据；抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量；将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据；抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤后得到真实

属性数据，获取所述真实属性数据的属性值；将所述实体数据矩阵中的元素和所述真实属性数据的属性值入参到可信度识别模型，出参后得到所述知识数据的可信度，将所述可信度与预设的可信度阈值进行比较，若大于所述可信度阈值则将抽取出的所述原始属性数据融合，否则不融合。

5 17、根据权利要求 16 所述的一种存储有计算机可读指令的存储介质，其中，所述获取知识数据来源中的数个知识数据时，使得所述处理器执行以下步骤：

发送知识数据抽取指令至待抽取知识数据的所述知识数据来源；接收所述知识数据来源的反馈信息，从所述反馈信息中抽取出数据来源类型的关键词，根据所述关键词确定所述知识数据来源的类型；获取所述知识数据来源的类型
10 对应的抽取方法，根据所述抽取方法抽取所述知识数据来源的数个知识数据。

18、根据权利要求 16 所述的一种存储有计算机可读指令的存储介质，其中，所述抽取任一所述知识数据中的实体数据，将所述实体数据进行向量化转换，生成多维词向量时，使得所述处理器执行以下步骤：

设置一抽取所述知识数据中的实体数据的初始语段，所述初始语段中至少
15 包含一个所述实体数据；根据所述初始语段的语段长度，将所述知识数据分割成数个初始子数据块，若任意一个所述初始子数据块中包含有两个或者以上的实体数据，则将所述初始子数据块进行再次分割得到只包含一个所述实体数据的最终子数据块；抽取所述最终子数据块中的实体数据，提取所述最终子数据块中的实体数据的语义特征，应用词向量转化法将所述语义特征转换成初始多
20 维词向量，将所述最终子数据块的语段长度作为系数与所述初始多维词向量进行乘积后得到最终多维词向量。

19、根据权利要求 16 所述的一种存储有计算机可读指令的存储介质，其中，所述将所述多维词向量进行降维后得到二维词向量，将所述二维词向量转置后与原所述二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向
25 量化的实体数据时，使得所述处理器执行以下步骤：

获取所述多维词向量中的每一个样本点的 K 个近邻点；根据每个样本点的 K 个近邻点，建立所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ；根据所述每个样本点的局部权重矩阵 $W_i = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$ ，将每个样本点映射到低维空间，映射后得到二维词向量 $Y = \{y_1, y_2, \dots, y_N\}$ ；将所述二维词向量转置后得到转置二维
30 词向量，将所述二维词向量与所述转置二维词向量乘积得到实体数据矩阵，所述实体数据矩阵中的元素为向量化的实体数据。

20、根据权利要求 16 所述的一种存储有计算机可读指令的存储介质，其中，所述抽取任一所述知识数据中的原始属性数据，对所述原始属性数据进行过滤

后得到真实属性数据，获取所述真实属性数据的属性值时，使得所述处理器执行以下步骤：

抽取所述任一所述知识数据中的原始属性数据，离散化处理所述原始属性数据后得到所述原始属性数据的离散值；根据所述知识数据中的所述原始属性数据的数量，获取所述原始属性数据对应的向量维度；将所述离散值和所述向量维度进行做差，若差值在预设的误差阈值以内，则所述原始属性数据为真实属性数据，若所述差值不在所述误差阈值以内，则根据所述差值去除所述原始属性数据中的多余属性数据，得到所述真实属性数据；根据所述真实属性数据的数量，获取所述真实属性数据对应的向量维度，建立真实属性数据向量；将所述真实属性数据向量降维后形成一真实属性数据矩阵，获取所述真实属性数据矩阵的特征值，所述特征值为所述属性值。

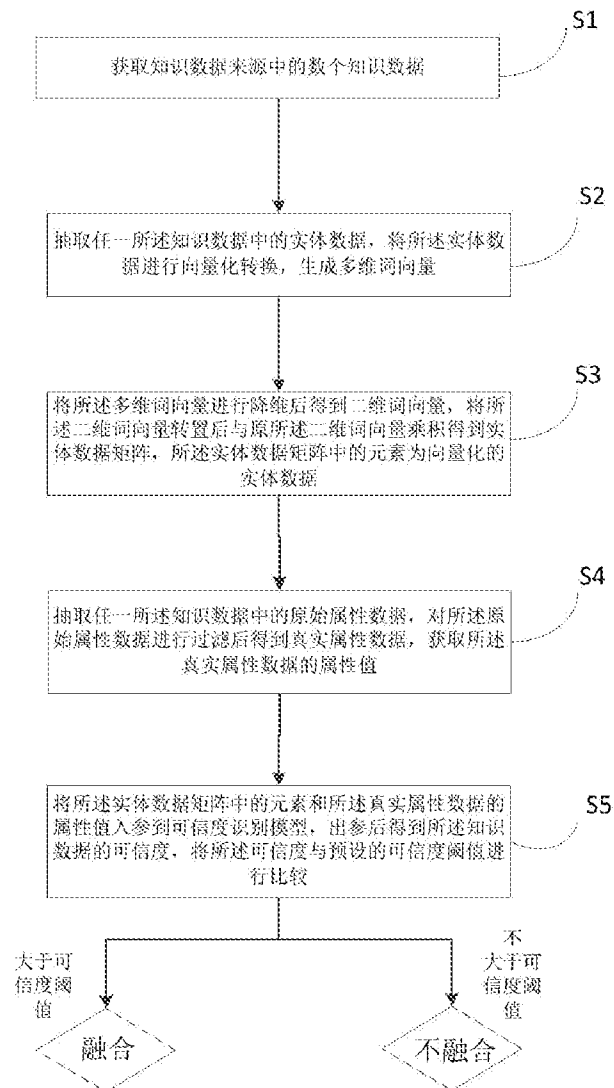


图 1

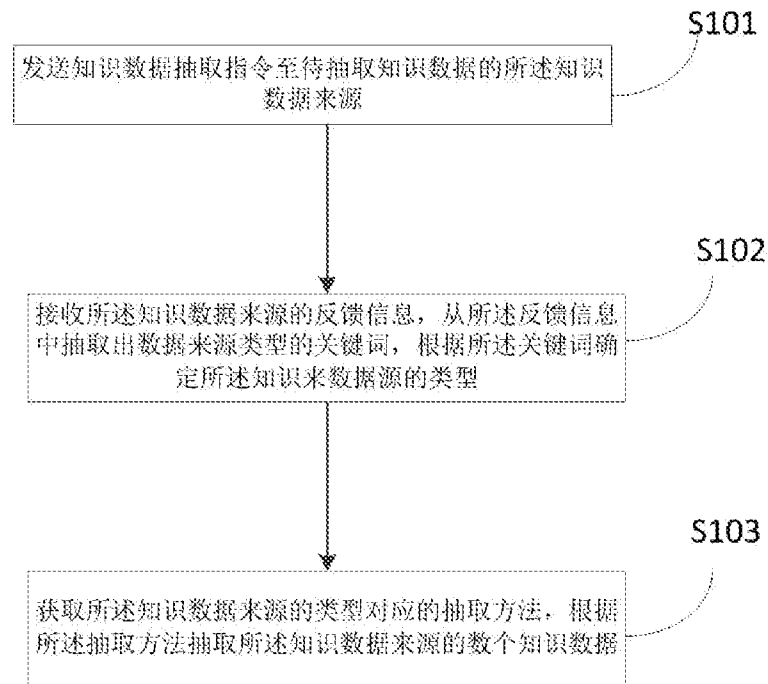


图 2

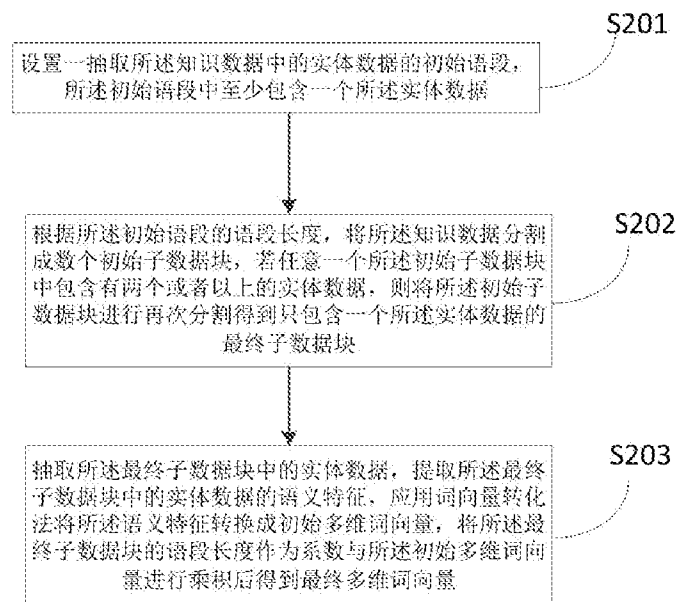


图 3

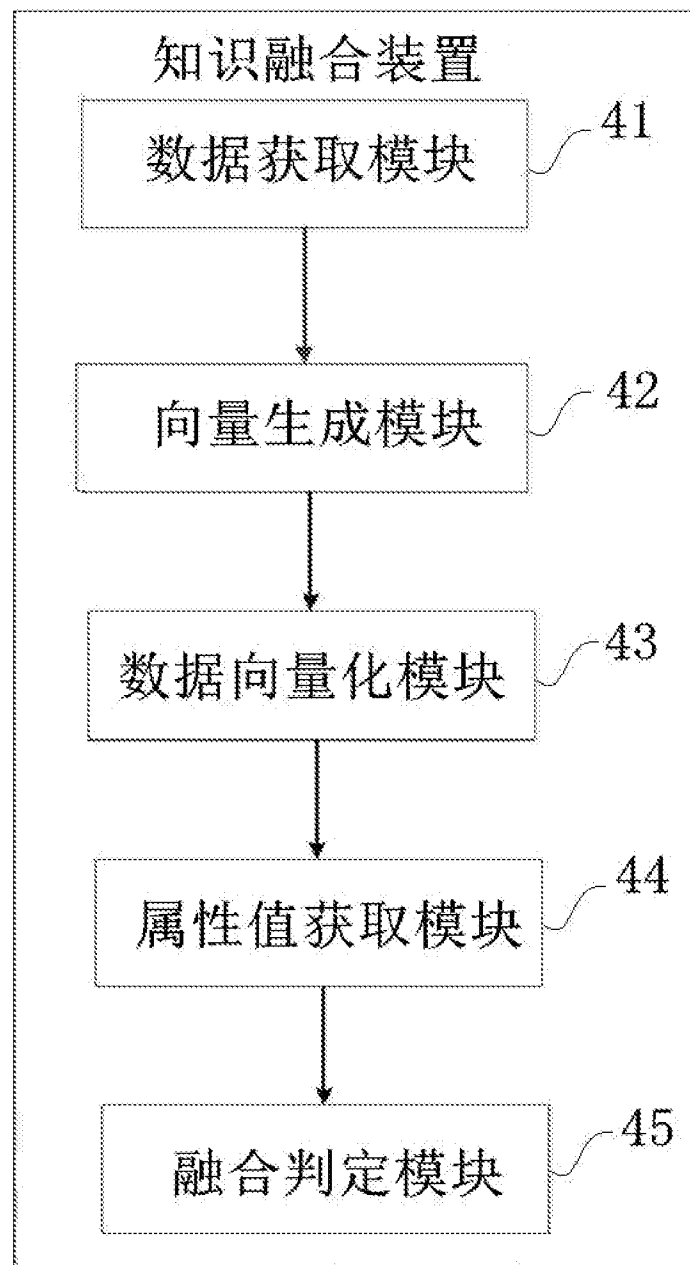


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/092597

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i; G06F 16/951(2019.01)i; G06F 16/36(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K; G06F; G06N; G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNKI, SIPOABS, DWPI: 知识, 融合, 数据, 源, 实体, 属性, 向量, knowledge, fusion, data, source, entity, attribute, vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109886294 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 June 2019 (2019-06-14) entire document	1-20
A	CN 108647318 A (BEIHANG UNIVERSITY et al.) 12 October 2018 (2018-10-12) entire document	1-20
A	CN 108536664 A (EAST CHINA NORMAL UNIVERSITY) 14 September 2018 (2018-09-14) entire document	1-20
A	US 2018268024 A1 (IBM) 20 September 2018 (2018-09-20) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

23 September 2019

Date of mailing of the international search report

09 October 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/092597

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109886294	A	14 June 2019	None	
CN	108647318	A	12 October 2018	None	
CN	108536664	A	14 September 2018	None	
US	2018268024	A1	20 September 2018	None	

国际检索报告

国际申请号

PCT/CN2019/092597

A. 主题的分类

G06K 9/62(2006.01)i; G06F 16/951(2019.01)i; G06F 16/36(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K; G06F; G06N; G06Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNKI, SIPOABS, DWPI:知识, 融合, 数据, 源, 实体, 属性, 向量, knowledge, fusion, data, source, entity, attribute, vector

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109886294 A (平安科技深圳有限公司) 2019年 6月 14日 (2019 - 06 - 14) 全文	1-20
A	CN 108647318 A (北京航空航天大学 等) 2018年 10月 12日 (2018 - 10 - 12) 全文	1-20
A	CN 108536664 A (华东师范大学) 2018年 9月 14日 (2018 - 09 - 14) 全文	1-20
A	US 2018268024 A1 (IBM) 2018年 9月 20日 (2018 - 09 - 20) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 9月 23日

国际检索报告邮寄日期

2019年 10月 9日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

张一良

电话号码 62089422

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/092597

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109886294	A	2019年 6月 14日	无	
CN	108647318	A	2018年 10月 12日	无	
CN	108536664	A	2018年 9月 14日	无	
US	2018268024	A1	2018年 9月 20日	无	