

工业设备维修知识图谱构建与应用

瞿智豪¹, 胡建鹏¹, 余思源², 黄子麒¹, 王泉雄¹

¹上海工程技术大学, 上海

²通用电气软件(上海)有限公司, 上海

收稿日期: 2021年9月24日; 录用日期: 2021年10月20日; 发布日期: 2021年10月28日

摘要

在工业智能化革命快速发展的助推之下, 知识图谱正越来越多地应用到工业领域中。本文以某散热膜生产工厂生产设备维修记录为数据源, 利用自顶向下和自底向上相结合的构建方式, 首先对原始数据进行处理与分析, 在专业人员的指导下, 完成对知识图谱模式层(本体)的构建, 再通过对数据源的抽取, 按照模式层的本体间关系, 构建知识图谱三元组; 最后, 通过Neo4j图数据库将其储存起来, 并以节点和边的图谱方式直观地将所构建的工业设备维修知识图谱表示出来。本文所构建的工业设备维修知识图谱可应用于维修人员评价系统、智能问答系统, 智能推荐系统以及维修异常预警系统。

关键词

工业智能化, 知识图谱, 知识抽取, 本体, 图数据库

Construction and Application of Industrial Equipment Maintenance Knowledge Graph

Zhihao Qu, Jianpeng Hu, Siyuan Yu, Ziqi Huang, Xiaoxiong Wang

¹School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai

²GE Software Shanghai Co., LTD., Shanghai

Received: Sep. 24th, 2021; accepted: Oct. 20th, 2021; published: Oct. 28th, 2021

Abstract

Driven by the rapid development of industrial intelligent revolution, knowledge graph is increasingly applied to the industrial field. In this paper, we take the maintenance records of production equipment in a heat dissipation film production factory as the data source, and we use the construction method of the combination of top-down and bottom-up. Firstly, we process and analyze the original data, and complete the construction of the knowledge graph pattern layer (ontology)

文章引用: 瞿智豪, 胡建鹏, 余思源, 黄子麒, 王泉雄. 工业设备维修知识图谱构建与应用[J]. 软件工程与应用, 2021, 10(5): 644-653. DOI: 10.12677/sea.2021.105069

with the guidance of professionals. Secondly, through the extraction of data sources, according to the relationship between ontology in the pattern layer we construct the triples of knowledge graph; finally, the knowledge graph we construct is stored in the graph database called Neo4j. The constructed industrial equipment maintenance knowledge graph is intuitively expressed in the form of nodes and edges. This knowledge graph can be applied to build the evaluation system of maintenance personnel, intelligent question and answer system, intelligent recommendation system and maintenance anomaly early warning system.

Keywords

Industrial Intelligence, Knowledge Graph, Knowledge Extraction, Ontology Construction, Graph Database

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在工业 4.0 和中国制造 2025 等政策以及新兴的人工智能技术的推动下, 工业领域正在经历着一场史无前例的智能化革命。

知识图谱(Knowledge Graph, KG)是谷歌公司在 2012 年正式提出的。本质上说, 知识图谱就是表示实体之间关系的语义网络。人工智能技术的一个重要分支, 通用知识图谱(GKG)已经在智能问答、智能搜索、智能推荐等方面表现出其强大的能力。随着知识图谱相应技术的不断发展和各领域各行业对实现其智能化的迫切需求, 深度更深、粒度更细、专业度更高的领域知识图谱(DKG)成为现今知识图谱研究和应用的主要方向, 在医疗、金融、教育等领域, DKG 已经取得了一些应用[1]。

工业是国民经济的血液, 产出了除食物原料外一切人类生存所需物质条件。传统意义上, 工业制造需要耗费大量的人力物力进行生产, 这样不仅耗时耗力, 而且在如采矿业、石油化工业等危险性行业中, 容易发生重大安全事故。因此, 工业领域进行智能化的需求显得尤为重要。如今, 在工业生产上, 我国工业部门已经实现了相当程度的自动化, 在一些行业甚至实现了无人化生产, 但是在设备检测, 维修等方面仍主要依赖于质检人员和维修人员的主观判断和经验。在一些信息化程度较好的工厂内, 一般会用 MES(制造执行系统)提升管理水平, 但利用 MES 指导或辅助人员完成维修工作只能以关键词形式查询到某一次维修记录, 无法体现某一次(类)故障与其他故障的关联性, 不能或不能很好地协助人员完成维修与质检任务。因此, 将结构化、半结构化与非结构化故障检测、维修处置文本等知识转变成直观地、可视的、结构化的形式是工业生产亟需的。

目前, 知识图谱在工业领域上的应用在国内外都有一些探索。百度的工业大脑解决方案基于知识图谱构建技术、工业领域知识、企业数据和经验等帮企业实现对多方知识的融合和深度关联分析, 应用于设备检修、故障预测和生产预测等场景。达观数据的工业知识图谱平台可以从输入的文本中自动构建图谱或根据给定的模式手动构建; 文献[2] [3] [4], 在飞机电源系统故障诊断、电网故障处置和煤矿装备维修领域构建了图谱, 可用于相关故障和维修的问答与建议。

本文利用在某散热膜生产工厂的设备维修记录, 以及互联网上获取的相关文本为数据源, 根据该工厂实际构建了本体层与图谱模式, 通过对数据的挖掘与处理, 将处理好的结构化数据利用图数据库 Neo4j

转换为图谱形式，初步实现可以用于搜索与辅助维修的知识图谱。

2. 知识图谱构建流程

知识图谱的构建，一般分为自底向上和自顶向下两种[5]。自底向上指通过对数据层的处理，包括对非结构化数据进行实体和关系抽取，对数据归纳总结后，形成模式层，并不断优化结构；自顶向下则是预先定义好模式层，并依次对数据进行知识抽取，并不断优化模式层；此外还可以将自顶向下和自底向上两种方法结合起来的混合法。本文的工厂设备维修场景具有一定的专业性，如果采用自顶向下的方式需要具有一定的专家经验指导，而使用自底向上的方法又需要大量的数据用以归纳出模式层，因此，本文使用自顶向下和自底向上结合的混合法构建工业设备故障及维修知识。

本文知识图谱的构建流程是首先对取得的某生产散热膜的工厂的一定时间段内的维修记录进行处理，由于涉及到工作人员姓名和设备厂商等信息，需要对一些条目进行脱敏处理，然后要对表中数据进行一些处理；其次，在该工厂管理人员的指导下，不断修改完善本图谱模式层即领域本体层；提取出所获数据的实体与关系，最后将实体与关系、属性等，按已构建的本体关系链接到图数据库 Neo4j 储存并显示为有向图的形式。图谱构建流程如图 1 所示。

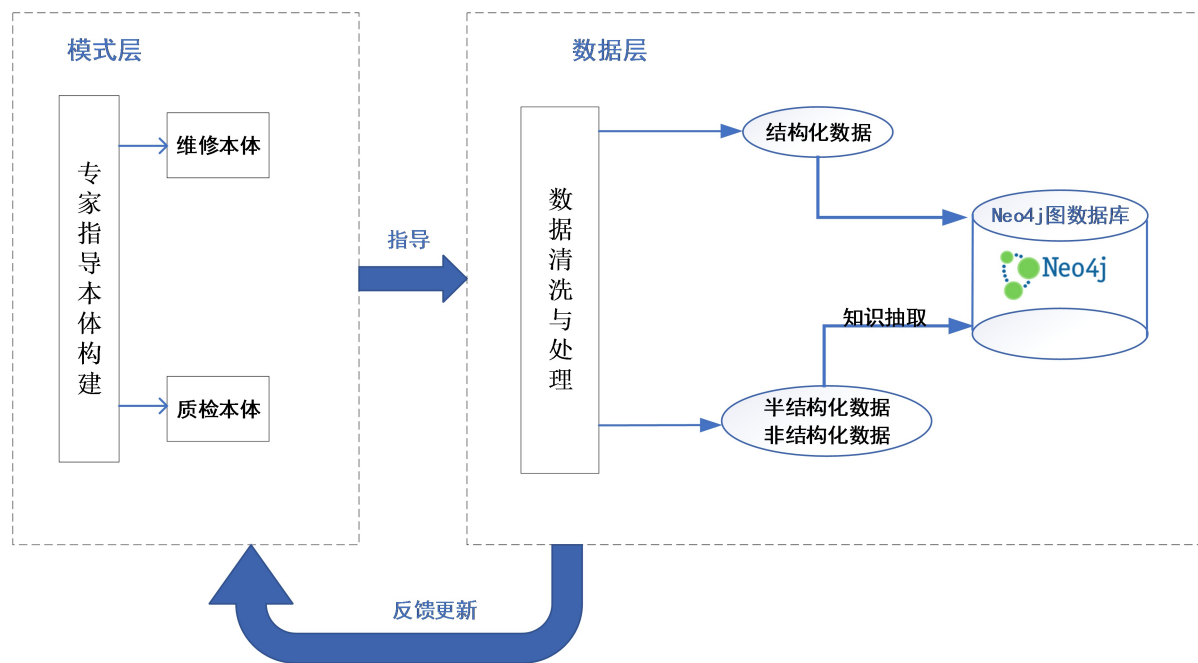


Figure 1. The construction process of the knowledge graph
图 1. 知识图谱构建流程

3. 数据处理与分析

3.1. 数据处理

本文的数据来源于某散热膜生产商，主要内容为其所记录的设备维修记录，以及用以解释记录内容的字典、设备对照表、检验项目对照表和车间数据等。设备维修记录是本文所构建图谱的数据来源，主要是结构化和半结构化数据。由于维修记录是由当值工作人员进行录入与管理的，虽然有一定的规范但原始数据仍有一些影响后续工作的内容如空记录、不正常数值错误(如同一检测标准表中数值不同)等。对于这些数据，如表 1 的这种空记录，可以结合前面的检测数据判定是否合格的记录，主体内容与正常文

本一致，是有价值的可以保留的。

Table 1. Empty records No.1

表 1. 空记录 1

项目结果	检验结论
合格	合格
合格	
合格	
合格	(空)
合格	
合格	
(空)	合格

而表 2 所示空记录作为维修记录关键信息的维修方式和人员没有记录，且无办法将其补充，所以需要将其视为无效记录，以删除处理。

Table 2. Empty records No.2

表 2. 空记录 2

维修方式	维修人员
(空)	
更换固态继电器	XXX
重新接线	XXX

如表 3 所示是一种不正常数值错误，即同一检测项目但规定检测值上下限有区别，这种数据可以根据其他批次的相关内容进行校正。

Table 3. Value errors

表 3. 数值错误

检验项目	检验类型	上限值	下限值
凹凸点&气泡	定量	20	0
凹凸点&气泡	定量	7	0

3.2. 数据分析

构建本体层，必须对所拥有的数据进行分析，详细了解所处理数据的内容，想要构建好本文所使用的石墨散热膜生产工厂的

设备维修、产品质检知识图谱，必须要了解该工厂的生产工序、检测逻辑等等内容。经过调研，该工厂即该产品的生产工序如图 2 所示。生产石墨散热膜需要经历分切车间将原料分切、包装并进入烧制压延车间，经高温炉、碳化炉等设备烧制、反应成所需石墨散热膜原版之后通过压延机将模板压制到所需的厚度，之后进入模切车间按照订单要求分切，最后进入成品车间进行产品质量检测，合格后封装成成品再统一包装。

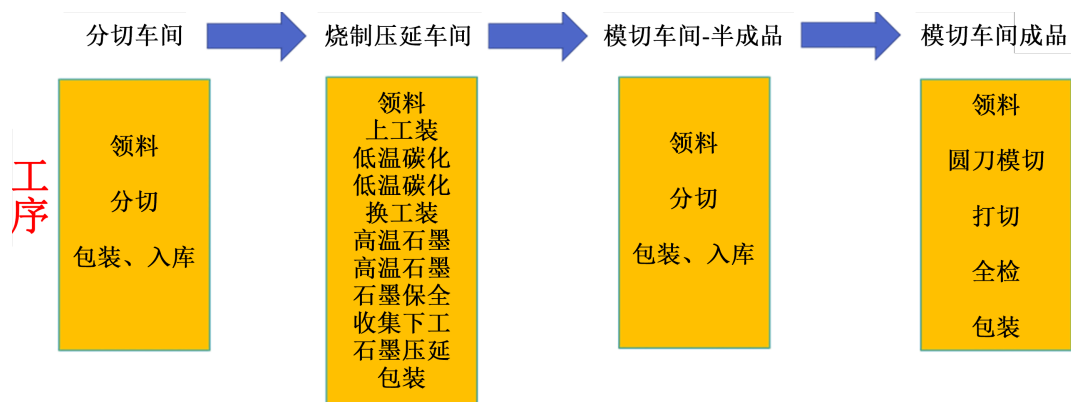


Figure 2. Production process of graphite heat dissipation film
图 2. 石墨散热膜生产工序

通过分析生产工序可以知道，在生产石墨散热膜的过程中，对产品质量起到决定性作用的同时也是使用最为频繁的设备主要是分切车间和模切车间的分切机、模切机，烧制压延车间的高温炉、碳化炉、压延机等。而通过对数据的分析，在设备维修场景中，维修次数和个体最多的设备如表 4 所示。

Table 4. Statistics of equipment maintenance times
表 4. 设备维修次数统计

设备	故障次数
碳化炉	2178
高温炉	345
压延机	170
模压机	82
倒卷机	38

可以发现，使用频率和数量较多的设备即上文所述高温炉、碳化炉、压延机等出现故障的次数是较多的。在后面的工作中要作为重点考量。

在维修场景中，设备维修的方式、好坏都取决于实施维修的技术人员，通过对维修人员的维修频次、维修设备种类等信息评价该人员能力，在后续工作中可以建立专家评分机制，赋予不同维修人员不同的评分，以供知识图谱下游任务(专家系统推荐)使用，也可以作为专家指导知识图谱本体层的构建。

此外，针对设备生产厂商生产的设备的故障率、某设备故障是否影响生产、某设备故障的类型(电气故障、机械故障、软件故障)及其对生产的影响、维修中有无备件更换对生产的影响等进行了统计分析，为后续建立维修图谱后的应用提供数据帮助。

4. 本体层构建

4.1. 本体

本体实际上就是对特定领域之中某套概念及其相互之间关系的形式化表达(formal representation) [6]。是针对概念与概念之间的联系以形式化的方式给出的明确定义[7]。本体构建是自顶向下构建知识图谱的过程中，构建顶层模式层的核心。本体包含概念、关系、属性三个核心要素：概念即对现实世界中事物的可观描述；关系即概念与概念间、概念与属性间的关系；属性即对概念的属性的描述。此外还有公理：现实世界中的真命题和实体：概念在现实世界中的个体的呈现。

Karp.P.D [8]等提出了构建领域本体的规则：

- 1) 明确性：本体刻画领域概念及概念间关系准确，不存在歧义。
 - 2) 完整性：对于概念、关系的定义应当完整。所给出的定义是完整的，完全能表达所描述领域术语的含义。
 - 3) 一致性：概念之间不存在冲突。由术语得出的推论与术语本身的含义是相容的，不会产生矛盾。
- 通过以上的数据分析和专业人士的指导，本文构建了石墨散热膜生产设备维修本体。

4.2. 设备维修本体构建

在本文所用的维修数据中，包括结构化数据和半、非结构化数据，其中结构化数据有十几项不同的内容，包括设备编号、维修人员、维修开始结束时间、故障类型、维修方式、维修完成情况等。经过分析和工厂专业人员的指导，这其中既有本体，又有属性。而图 3 为构建的设备维修场景本体间和本体与属性间关系图，其中本体概念用黄色标出，关系为蓝色箭头，属性为蓝色中空圆框。为方便理解，本文直接将本体显示为有向图形式。

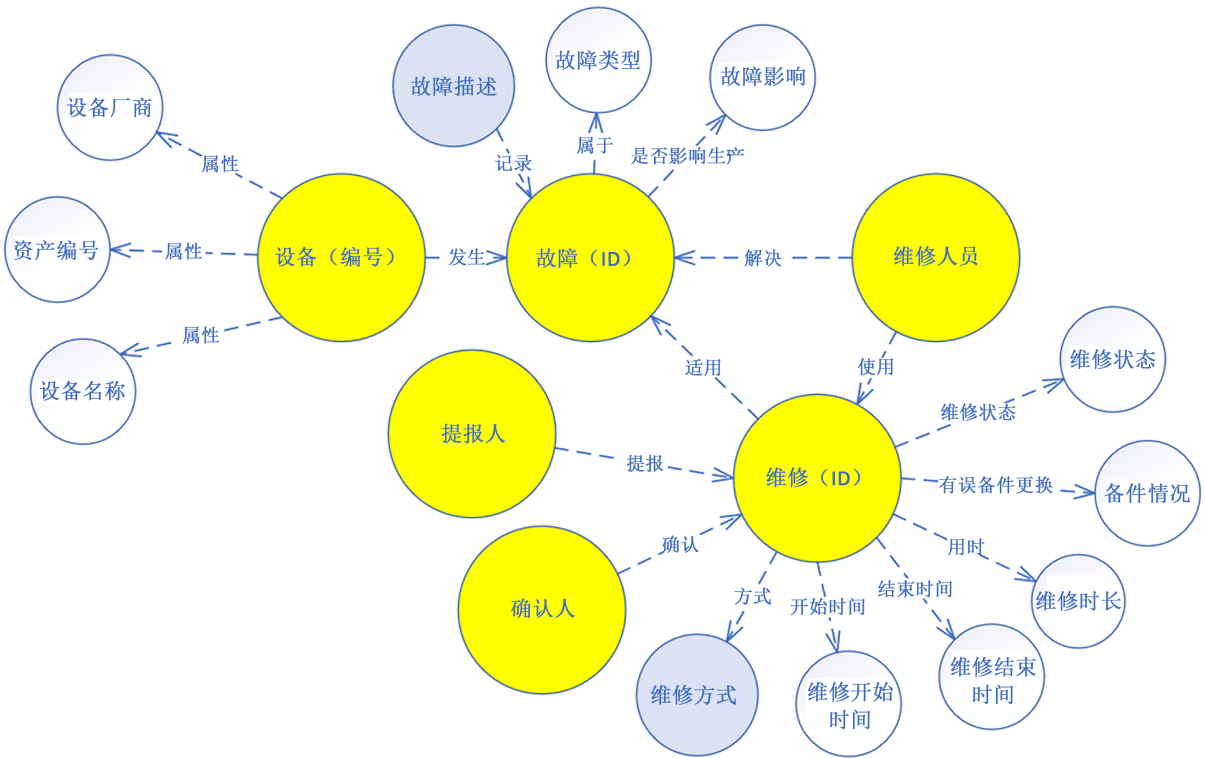


Figure 3. Ontologies of equipment maintenance
图 3. 设备维修本体

对于设备维修场景，各本体实体如表 5 所示：

Table 5. Entities of ontologies in the scene of equipment maintenance
表 5. 设备维修场景各本体实体

本体	实体
设备	设备编号

Continued

故障	故障 ID
维修人员	维修员姓名
维修	维修 ID
提报人	提报人姓名
确认人	确认人姓名

除了对应各本体的实体，本体间的关系的定义对于模式层的建立并引导图谱的构建也是至关重要的。
表 6 所示是设备维修场景和质检场景中各本体间的关系。表中本体列中两个本体与关系，如“设备 - 故障 - 发生”表示某设备发生了某故障。

Table 6. Relations between ontologies in the scene of equipment maintenance

表 6. 设备维修场景本体间关系

本体	关系
设备 - 故障	发生
维修人员 - 故障	解决
维修 - 故障	适用
提报人 - 维修	提报
维修人员 - 维修	使用
确认人 - 维修	确认

5. 数据层构建

在经过数据分析与专家指导构建好知识图谱的模式层之后，本文基于模式层，对知识图谱数据层进行了构建。

5.1. 结构化数据构建

根据模式层中，本体间的关系和本体属性的关系，可以直接将结构化数据以三元组的形式储存在数据库中，我们的结构化数据主要为人员与维修设备、质检产品间关系等的三元组形式如{维修员 XXX，维修，设备 XXX}；{质检员 XXX，检测，质检项目 XXX}；{维修 ID，有/无，备件}。

5.2. 半结构化、非结构化数据构建

非结构化数据指符合语言规范的文本等数据，在工业领域，一般包括设备维修记录文本、维修手册、质检报告等。而半结构化数据指具有一定结构性的文本数据，可以通过制定一定的规则将其转化为结构化数据，或将半结构化数据的非结构化部分按非结构化数据处理。

在模式层制定的本体属性关系的基础上，对非结构化数据，需要对其进行知识抽取，从而将其转化为三元组的形式。知识抽取包括实体抽取(命名实体识别)和关系抽取。

命名实体识别是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等。命名实体识别的方法分为三大类：1) 基于规则的命名实体识别指通过领域专家对文本的分析，构建一种模板，再按照模板与字符串相匹配的方式进行命名实体识别；2) 基于统计机器学习的方法利用机器学习方法进行命名实体识别，能较好地移植到其他语料。主要有隐马尔可夫链、决策树、条件随机场(CRF)、

支持向量机(SVM)等方法; 基于深度学习的命名实体识别[9]。近年来, 许多研究者将在图像处理等方面取得重大成果的深度学习的方法应用到自然语言处理的问题中, 取得的很多成果, 许多方法都超越了之前的传统算法。如卷积神经网络(CNN)、长短时记忆网络(LSTM)以及它们的改进等方法[10] [11]。

关系抽取是在命名实体识别的基础之上, 对文本中包含的实体间的关系以及关系对应的实体辨别出来的任务。主要有基于模式匹配的方法、基于有监督学习/半监督学习方法、无监督学习方法和基于表示学习方法等[12]。

本文数据除了结构化数据外, 主要是有一定自然语言文本的半结构化数据, 由于本文数据语料不长, 所以采用制定的规则的方法, 对文本进行知识抽取。通过工厂专家制定的规则: XXX 故障/漏水/漏气/异常/断电……导致/致使/损坏……XXX。对文本分词后, 得到本文所需的实体与关系。如图 4 所示为对文本分词结果的统计。

部件		动作	
水管-n	309	更换-v	2684
热电偶-n	235	处理-v	768
丙酮-nz	166	重新-a	336
螺丝-n	140	检查-vn	233
密封圈-n	137	疏通-v	185
开关-v	116	调整-vn	179
蝶阀-n	115	加热-v	179
可控硅-nz	102	紧固-a	172
真空泵-n	94	清洗-v	159
电机-n	92	连接-v	154
压力表-n	86	清理-a	145
保压阀-n	86	修复-v	123

Figure 4. Statistics of segmentation result

图 4. 分词结果统计

由图 4 所示统计数据和对分词后的语料进行分析, 本文数据大多符合事先制定的规则, 少部分语料使用人工抽取。

半结构化数据知识抽取后, 我们同样构造了相应的三元组如{XXX 短路, 致使, XXX 断电}等。

6. 知识图谱可视化

构建好数据层后, 需要将知识储存起来, 一般采用关系型数据库、RDF 三元组和图数据库形式储存知识图谱。为了能更直观得将知识图谱呈现出来, 比如实体关系三元组就是以节点和边构成, 且储存与查询都十分方便。本文使用图数据库 Neo4j 储存知识图谱。

Neo4j 采用与 SQL 语句类似的 CQL 语句对节点、关系、属性等进行储存与查询。如创建维修人员节点:

```
Create (n:维修人员 { name: 'XX' }) return n;
```

创建关系:

```
MATCH (a: 维修人员), (b:设备)
```

```
WHERE a.name = 'XX' AND b.name = '分切机 2'
```

```
CREATE (a)-[r:维修 {roles:['维修']}]>(b)
```

```
RETURN r;
```

也可以通过 Python 的 py2neo 工具包, 通过编写 python 程序, 创建节点与关系并导入到 Neo4j 中。本文根据构建的本体属性关系, 使用 py2neo 在 Neo4j 中构建出设备维修知识图谱如图 5 所示。

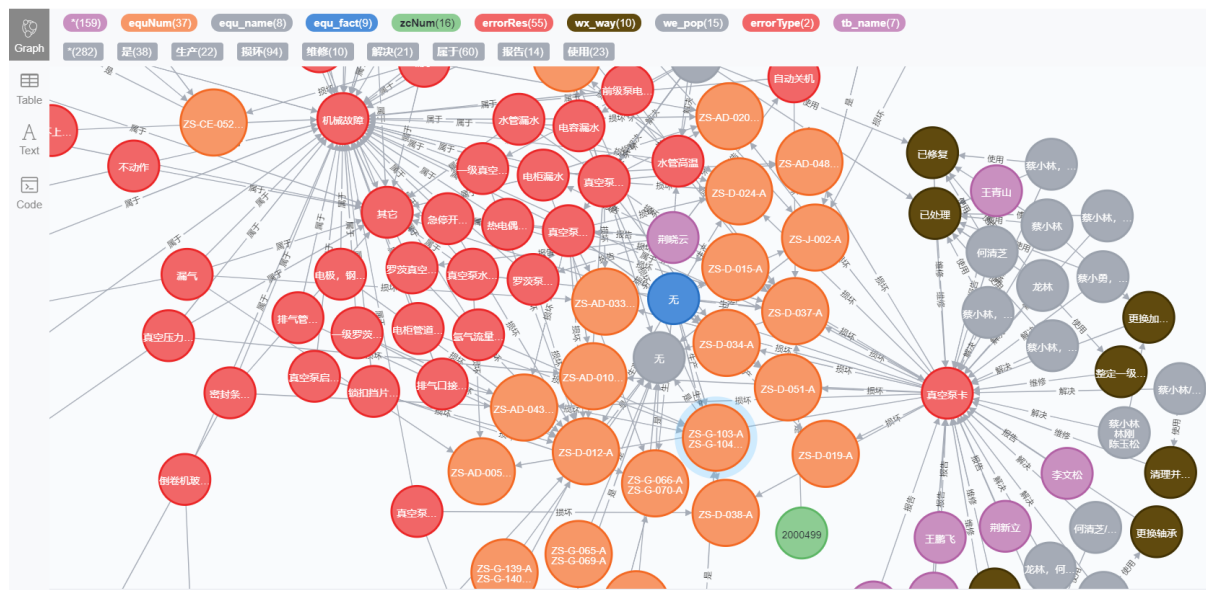


Figure 5. Presentation of knowledge graph
图 5. 知识图谱呈现

7. 工业设备维修知识图谱应用

在通用领域，知识图谱已经有了很多应用，而在专业领域，如医疗[13]，金融等领域，也已经实现了诸如医疗问题自动问答、金融问题自动问答，理财产品自动推荐等功能。

而在工业领域，由于工业部门的多样性和专业性，很难形成通用整个工业领域的知识图谱，而只能限定在某场景，比如本文所构建的设备维修场景和产品质检场景。而生产不同的产品又涉及到可能截然不同的生产设备和质检流程。一直以来，由于工业领域知识图谱的专业领域限制，一直没有较为通用的应用场景。

利用本文所构建的知识图谱，可以应用于以下几个系统：

- 1) 维修人员评价系统。利用知识图谱帮助专家直观地评价某工作人员的工作能力，判断维修人员精通的设备与类型，此外，还能通过链接某人员在一次维修操作后的连锁反应，更直观公正地评价维修人员的工作能力。
- 2) 建立智能问答系统。通过对设备部件或者损坏方式的解析，系统匹配到相应的记录，在图数据库中以节点和边的形式呈现，或直接以答案的形式呈现，以协助工作人员快速吸取以往经验，或帮助新人快速找到行之有效的方法。
- 3) 建立智能推荐系统。通过工作人员输入的相关信息，通过连接预测等技术，联想到工作人员可能要解决的问题，辅助其进行决策。或者根据工作描述的问题，为其推荐精通或擅长于此的技术人员。
- 4) 维修异常预警系统。建立设备维修异常预警系统，当维修人员的操作与知识图谱中关联的记录有所区别或以往相关操作造成维修失败或异常的情况时，系统给予维修人员预警并告知可能发生的状况，以此减少维修方法错误而浪费资源的情况发生。

8. 展望

如今，工业智能化正在如火如荼地进行着，知识图谱的应用为工业智能化进程提供了强劲动力。但由于工业领域分支多，专业性强，专业壁垒高的特点，始终无法实现如通用领域般体量巨大且高效的应用。

本文所创建的工业设备维修知识图谱可以帮助专业人士更好更高效地进行工作，也能为非专业人士提供专业的知识指导。

本文所述知识图谱应用场景是诸多工业领域纵向场景的一环。但是，在工业领域，涉及维修的文字内容除了本文所用的维修记录这种结构性较好的内容外，还有很多如维修日志、手册，乃至图表之类结构性较差的文本，在知识抽取时有诸多的问题，现在一些先进方法仍需要不断完善。同时，知识图谱在工业领域的应用仍然受困于通用领域所延伸出的应用方向，只有更多人投入到知识图谱在工业领域的应用研究，知识图谱才能更好地助推工业智能化的发展。

参考文献

- [1] 朱超宇, 刘雷. 基于知识图谱的医学决策支持应用综述[J]. 数据分析与知识发现, 2020, 4(12): 26-32.
- [2] 聂同攀, 曾继炎, 程玉杰, 马梁. 面向飞机电源系统故障诊断的知识图谱构建技术及应用[J/OL]. 航空学报: 1-19, <http://kns.cnki.net/kcms/detail/11.1929.V.20210825.1351.004.html>, 2021-09-23.
- [3] 郭榕, 杨群, 刘绍翰, 李伟, 袁鑫, 黄香鸿. 电网故障处置知识图谱构建研究与应用[J]. 电网技术, 2021, 45(6): 2092-2100.
- [4] 曹现刚, 张梦园, 雷卓, 段欣宇, 陈瑞昊. 煤矿装备维护知识图谱构建及应用[J]. 工矿自动化, 2021, 47(3): 41-45.
- [5] 付雷杰, 曹岩, 白瑀, 冷杰武. 国内垂直领域知识图谱发展现状与展望[J/OL]. 计算机应用研究: 1-15, <https://doi.org/10.19734/j.issn.1001-3695.2021.04.0095>, 2021-09-23.
- [6] 百度百科本体. <https://baike.baidu.com/item/%E6%9C%AC%E4%BD%93/17330?fr=aladdin>
- [7] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
- [8] Karp, P.D. and Gruber, T. (1994) A Generic Knowledge-Base Access Protocol. *Technical Report*.
- [9] 焦凯楠, 李欣, 朱容辰. 中文领域命名实体识别综述[J/OL]. 计算机工程与应用: 1-16, <http://kns.cnki.net/kcms/detail/11.2127.tp.20210526.1823.008.html>, 2021-09-23.
- [10] Collobert, R., Weston, J., Bottou, L., et al. (2011) Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, **12**, 2493-2537.
- [11] Chiu, J.P.C. and Nichols, E. (2015) Named Entity Recognition with Bidirectional LSTM-CNN. *Transactions of the Association for Computational Linguistics*, **4**, 357-370. https://doi.org/10.1162/tacl_a_00104
- [12] 刘辉, 江千军, 桂前进, 张祺, 王梓豫, 王磊, 王京景. 实体关系抽取技术研究进展综述[J]. 计算机应用研究, 2020, 37(S2): 1-5.
- [13] 侯梦薇, 卫荣, 陆亮, 兰欣, 蔡宏伟. 知识图谱研究综述及其在医疗领域的应用[J]. 计算机研究与发展, 2018, 55(12): 2587-2599.