

《智能信息处理》课程考试

## 基于本体的武器装备知识图谱构建

智泽镭

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021 年 12 月 20 日

# 基于本体的武器装备知识图谱构建

智泽镭

(大连海事大学 信息科学技术学院, 大连 116026)

**摘要** 为解决武器装备领域数据来源分散、数据间缺乏良好关联、表示形式不一致、存在大量冗余等问题,提出了一种基于本体的知识图谱构建方法,以利用知识图谱实现武器装备知识的有效管理和智能信息搜索,进一步挖掘武器装备数据的潜在价值。本方法首先基于领域知识构建领域本体,而后对收集到的大量原始数据进行抽取,对其中的冗余数据进行融合,并采用三元组形式对得到的实体、关系、属性等数据进行表示,最后构建到知识图谱中。构建好的知识图谱可为原始数据的抽取、融合提供指导,形成知识图谱优化完善的良性循环。

**关键词** 武器装备; 知识图谱; 本体;

## Construction of knowledge map of main battle weapon equipment based on ontology

ZHI Ze-Rong

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

**Abstract** To solve the weapons and equipment field data source dispersion, the lack of a good correlation between data, inconsistent representation, some problems such as a large amount of redundancy, this paper proposes a knowledge map building method based on ontology, to utilize the knowledge map to realize the effective management of weapons and equipment knowledge and intelligent information searches, further excavate the potential value of the weapon equipment data. This method firstly constructs domain ontology based on domain knowledge, then extracts a large amount of collected original data, fuses the redundant data, and expresses the obtained data such as entities, relations and attributes in the form of triples, finally constructs the domain ontology into the knowledge graph. The constructed knowledge graph can provide guidance for extraction and fusion of original data and form a virtuous cycle of knowledge graph optimization.

**Keywords** Weapons and equipment; Knowledge graph; Ontology;

## 1 引言

知识图谱的概念最早可以追溯到 20 世纪 60 年代形成的“语义网络”理论,2012 年谷歌首次提出“知识图谱”技术并用于搜索引擎的智能化实现,它由互连的节点和边构成,以图的形式呈现节点及其之间的关系[1]知识图谱包括通用知识图谱和领域知识图谱两大类。通用知识图谱主要强调知识的广度,而领域知识图谱则面向不同的领域,基于不同行业的不同需求定制,往往需

要领域专家的经验支持和反复讨论。例如,国土资源、测绘、电力、网络安全、情报等领域的知识图谱研究。

近年来,军事领域对于武器装备的“数据化”“信息化”要求越来越高,作战形式也向“联合”方向发展,单一军兵种或单一形式的数据资料远远不能满足战场需求。如何展示武器装备数据之间的关系,对武器装备数据进行进一步的分析,全面考量各类武器装备对于军事行动的作用,支撑相关领域相关内容的快速分析和智能搜索,成为当前亟待

解决的问题。作者尝试提出一种武器装备知识图谱的构建方法,以支持主战武器装备数据的智能信息搜索,进一步发挥大量武器装备数据的潜在价值。

## 2 相关概念

### 2.1 本体

本体是对客观的事物以一种形式化的、客观的并且系统化的方式进行描述。本体由哲学领域发起,对现实世界的客观事物进行本质化的描述。它在哲学中的定义为“对世界上客观存在物的系统地描述”,是客观存在的一个系统的解释或说明,关心的是客观现实的抽象本质。本体可以在语义层次上描述知识,可以看成描述某个学科领域知识的一个通用概念模型。人们用许多不同的方式交流,通过说和听,做手势,或各种形式的文本。非结构化文本信息的存在使得有效地执行知识管理活动变得更加困难。计算机能够理解普通语言并与人类进行对话的想法是几个世纪以来的一个梦想。

### 2.2 知识图谱

知识图谱本身是一个具有属性的实体通过关系链接而成的网状知识库。知识图谱是结构化的语义知识库,用于以符号形式描述物理世界中的概念及其相互关系其基本组成单位是“实体—关系—实体”三元组,以及实体及其相关属性—值对,实体间通过关系相互联结,构成网状的知识结构[2]。从图的角度来看,知识图谱在本质上是一种概念网络,其中的节点表示物理世界的实(或概念),而实体间的各种语义关系则构成网络中的边。由此,知识图谱是对物理世界的一种符号表达。

### 2.3 关系

关系指事物之间相互作用、相互影响的状态,或人和人之间或任何事物之间某种性质的联系。在武器装备领域,本体及其之间的关系主要分为6类:上下关系主要是根据范围确定父概念与子概念,由于上下关系的数量有限,在构建时可主要依赖研究报告和热门军事网站中所描述的上下层次关系[3];同义关系主要指概念上相同或者相似的表达,主要从字面表示的不

同对同一个概念或者本体进行抽取;属性关系为武器装备及其特征之间的关系;整体与局部关系主要存在于武器装备的组成当中,本文研究内容的颗粒度为武器平台的级别,因此该类关系并未具体体现;概与实例关系主要是指上层概念与下级具体实例之间的关系;但并不是所有的关系都能归结到这5种关系上来,其他无法明确归类的关系在本文中统一称为一般关系。

### 2.4 关系描述

RDF是W3C推荐标准,是一种资源描述语言,它受到元数据标准、框架系统、面向对象语言等多方面的影响,用来描述各种网络资源,其为人们在WEB上发布结构化数据提供了一个标准的数据描述框架。RDF用二元关系模型来表示事物之间的语义关系,即使用三元组集合的方式来描述事物和关系。三元组也是知识图谱中知识表示的基本单位,可用于表示实体与实体之间的关系,或者实体的某个属性的属性值是什么。从内容上看三元组的结构为“资源-属性-属性值”的形式。根据抽取的大量数据,将武器装备的属性及武器装备之间的关系进行梳理总结。根据武器装备属于不同的层次、分类来确定本体之间存在什么关系;根据描述本体性质的内容,来提炼本体的属性,可以得到<轰炸机,研发国家,国家>这样的三元组,既是轰炸机和国家两个实体间的关系,又可以将国家作为描述轰炸机的属性之一。

## 3 武器装备领域知识图谱的构建

### 过程

#### 3.1 构建流程

本文在开始构建知识图谱时采取自上而下的构建方法,首先构建领域本体,而后对收集到的大量数据进行抽取,并对其中的冗余数据进行融合,形成符合要求的三元组形式,最后利用知识图谱工具构建知识图谱。构建好的知识图谱可用于指导数据的抽取过程,以不断丰富和完善知识图谱。

#### 3.2 本体设计

本文主要聚焦于平台级的武器装备进行分类。对于以某种平台为载体的装备,例

如专门用于通信保障的通信车、用于地理测绘的测绘车辆等, 统一归结到车辆类别中; 而对于以独立形态存在的通信电台等装备, 可以在领域本体中扩充相应的本体概念, 例如通信电子装备。以下是我们定义的部分武器装备领域本体概念, 主要围绕平台级武器装备展开[4]。依据上述扩充原则, 在此基础上可以扩展出其他类型的武器装备本体概念。

武器装备领域本体={飞行器、舰船舰艇、枪械与单兵武器、坦克装甲车辆、火炮、导弹武器、爆炸物、通信电子装备、软件类设备……};

飞行器={航空器、航天器};航空器={战斗机、攻击机、轰炸机、教练机、预警机、侦察机、反潜机、电子战机、无人机、运输机、飞艇、试验机、加油机.....};舰船舰艇={航空母舰、战列舰、巡洋舰、驱逐舰、护卫舰、两栖作战舰艇、核潜艇、常规潜艇、水雷战舰艇、导弹艇、巡逻舰、巡逻艇.....};枪械与单兵武器={非自动步枪、自动步枪、冲锋枪、狙击枪、手枪、机枪、霰弹枪、火箭筒、榴弹发射器、刀具.....};坦克装甲车辆={步兵战车、主战坦克、特种坦克、装甲运兵车、装甲侦察车、装甲指挥车、工程抢修车、布雷车、扫雷车.....};

火炮={榴弹炮、加农炮、加农榴弹炮、迫击炮、火箭炮、高射炮、坦克炮、反坦克炮、无后坐炮、装甲车载 炮、舰炮、航空炮、自行火炮、弹炮结合系统……} ;导弹武器={反弹道导弹、潜舰导弹、空舰导弹、岸舰导弹、舰导弹、空导弹、地空导弹、舰空导弹、地导弹、舰地导弹、空地导弹、潜地导弹、反辐射导弹、反坦克导弹……} ;

### 3.3 实体、属性选取

在对本体进行设计后,会形成本体概念的层次体系,体系中的本体都是经过验证的准确概念。在对大量数据进行处时如遇到不在体系中,但是符合本体条件的词语,可进行人工判定并加入体系。根据定义好的本体概念层次,结合获取的大量数据,对武器装备领域的知识进行实体抽取,主要以本体实例的形式充实到知识图谱中。同时,抽取本体的各种属性对知识图谱进行丰富,例如,

每类武器装备实例都有研发国家、生产日期、配发部队等[5]。下面以 B-2 轰炸机为例进行说明。基于上文建模的 6 类关系,对数据进行关系、属性抽取处理,得到<实体,关系,实体>、<实体,属性,属性值>等三元组。继续对上文的 B-2 轰炸机示例进行抽取可得到以下三元组:其中表示关系的三元组有:<B-2 轰炸机,携带,近距攻击导弹>;<B-2 轰炸机,携带,B83 炸弹>;<B-2 轰炸机,投入,科索沃战争>;<B-2 轰炸机,投入,伊拉克战场>;<B-2 轰炸机,投放,EGBU-28 制导炸弹>;<B-2 轰炸机,炸毁,通讯塔>;<通讯塔,位于,巴格达市>。表示属性的三元组有:<B-2 轰炸机,机长,21.03m>;<B-2 轰炸机,机高,5.18m>。

抽取得到的知识中可能存在大量的冗余,需要对冗余数据进行融合。同义词和近义词是典型的冗余数据之一,可作为同义关系来处理。将语义相似的、表达方式不同的词语定义为同义关系,再遇到此类词语时,将其按照同一类模板进行处理。另外,时间和地点数据也可以用于进行知识融合。通过判断时间和地点的一致性可以把不同实体有效关联起来。可以对时间和地点进行单独处理,以辅助武器装备领域知识图谱的构建,提高知识图谱的准确性。知识图谱还要随着数据的更新换代而不断迭代更新,因此需要经常对数据进行获取处理,并对图谱进行更新维护。随着武器装备的发展,其领域本体、关系等也需要适时更新。

### 3.4 知识图谱构建结果

图 1 是利用 Neo4j 图数据库构建的知识图谱。显示了航空器、B-2 轰炸机、科索沃战争等实体或实例及其之间的关系。

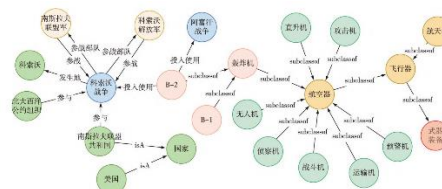


图 1 知识图谱示例

**总结** 本文阐述了武器装备领域知识图谱的构建方法，重点定义了其领域本体及关系，讨论了实体、关系、属性等抽取过程。构建

知识图谱时，首先要确保本体概念层次的完整性和准确性，再对下层数据进行扩充，这样才符合武器装备领域的特点。构建知识图谱时，经常面临的一个现实问题是，虽然数据量很大，但真正对构建图谱起作用的不多，而且武器装备领域的很多敏感资料是采集不到的。因此如何保证图谱的实时性和有效性是值得进一步研究的问题。

### 参 考 文 献

- [1]肖仰华. 知识图谱:概念与技术[M]. 北京: 电子工业出版社, 2020.
- [2]刘广禹,陈瑶瑶,王娇.基于 CiteSpace 的国土资源“一张图”知识图谱分析[J].国土资源信息化,2021(02):22-27.
- [3]路威,赵丽君.兵要知识图谱的构建与应用研究 [J/OL]. 测绘地理信息 :1-6[2021-12-20].<https://doi.org/10.14188/j.20956045.2020094>.
- [4]李代祎,盛杰,刘运星,马宗民.基于知识图谱的军事武器问答系统[J].指挥信息系统与技术,2020,11(05):5865.DOI:10.15908/j.cnki.cist.2020.05.011.
- [5]夏正洪,万健,朱新平.基于科学知识图谱的效能评估研究可视化分析[J].火力与指挥控制,2020,45(03):133-137.