

# CCKS 2022 技术评测任务书

## 外军无人系统知识图谱构建评测任务

无人系统已成为现代军事力量的重要组成，在各类军事行动中发挥着日趋重要的作用。军事垂直领域知识图谱的构建，受到数据、领域知识、构建技术三方面因素影响。当前，外军无人系统知识图谱构建在数据方面存在以下问题，一是缺乏权威性，数据来源以百科、新闻网页抓取为主，数据生产者与数据可信性之间没有约束关系；二是缺乏精准性，部分信源数据之间存在冲突、甚至明显错误；三是数据缺乏系统性，无人机之外的无人车、无人艇等其他无人系统涉及较少，技术组成、作战运用相关领域数据匮乏，对体系研究的支持能力差。基于此类数据构建的知识图谱，存在领域知识与构建技术无法弥补的“先天不足”，既“**难以保障严肃分析任务、难以服务实际业务系统、难以嵌入各类应用框架**”。为此，评测组织方联合中国知识图谱与语义计算大会（中国中文信息学会主办，国内知识图谱领域影响力最大、覆盖范围最广的学术会议，简称 CCKS）、红山开源平台（<https://osredm.com>），共同开展本次评测任务。**组织方**着重提供权威、精准数据以及部分领域知识指导，数据覆盖陆、海、空、天、潜等各类型无人系统，包含背景、作战、技术、性能等信息；**参赛队**专注图谱构建技术，对图谱构建过程中涉及的技术、方法、模型、训练数据等不作限制，自动化、半自动化以及其他混合方式均可，鼓励参赛队面向工程实际开展探索；**评测上**，重点关注图谱的任务能力，结合实际任务想定开展评测。本次评测任务完成后，组织方将补充数据及领域知识，构建军事垂直领域权威知识图谱，并依托红山开源平台，根据不同用户类型，提供线上图谱分析支持服务，表现突出的参赛队可参与相关图谱的维护、升级、标准化、使用等工作，并将获得 CCKS2022 组委会颁发的获奖证书及奖金。

### 1、任务定义

#### 输入：

1.1 原始数据。包含图片、表格以及非结构文本信息的 word 文档。

1.2 图谱任务想定。以文字形式描述希望从知识图谱中得到支撑的具体问题，用于图谱任务阶段评测。

#### 输出：

基于原始数据构建生成的知识图谱，包括实体列表、实体属性列表、实体间关系列表三份文件。

1.3 实体列表文件（entities.csv）。包含所构建知识图谱的所有实体，格式为“实体 ID、实体名称、实体标签”，要求“实体 ID”必须唯一，列表中的不同实体，只能有一个对应的“实体 ID”，不允许出现重复实体

ID; 多标签情况，标签之间用“;”隔离。

1.4 实体属性列表文件 (entities\_attr.csv)。包含所构建知识图谱所有实体的属性，格式为“实体 ID、实体属性、属性值”（实体 ID 取自实体列表文件）。

1.5 实体间关系列表文件 (relationships.csv)。包含所构建知识图谱的实体间关系，格式为“头实体 ID、尾实体 ID、关系属性/标签等”（实体 ID 取自实体列表文件）。

1.6 任务分析报告。包括对 1.2“任务想定”的解析过程（给知识图谱提供的输入）及结果，得到的知识图谱输出内容及拓扑结构（最好有可视化内容），对输出的分析。

1.7 图谱构建技术报告。清晰描述图谱构建过程，采用的技术、数据来源（文件或地址）及相关细节。

#### 数据示例：

##### 原始数据：

{

装备：MK 18 MOD 1 (剑鱼) 无人潜航器

用户：美国海军第一特种清除队 (NSCT ONE)

美国海军直属爆炸物和军火处理队 (EOD)

研制单位：Hydroid, LLC



**背景：**MK 18 MOD 1剑鱼无人潜航器是给NSCT ONE和EOD配备的工具箱中的一部分。它能够在极浅水域 (VSW) (10~40 FSW) 和通海航道区域进行低能见度的探测和侦察，为两栖登陆舰登陆，反水雷作战和水文绘图提供支援。其特点是，体积小 (两人便捷式)、造价低 (无意丢失也不会对部队造成灾难性的损失)，能够用于多用途平台。达到全面作战能力后，首批三套系统中的第二、第三套交付给了NSCT ONE。另外的系统将用于EOD进行中国大陆以外 (OCONUS) 的支援工作。该系统可以通过声频收发器在长基线或者超短基线模式下，或者通过P码GPS进行导航定位。上视和下视声学数字计程仪能够改进艇上惯性导航的精准度。舰载传感器包括水流混浊、水流温度和导热性测量仪、侧扫声呐以及下视摄像机。

#### 基本数据：

| MK 18 MOD 1 式 (剑鱼) 无人潜航器 |                     |      |                              |
|--------------------------|---------------------|------|------------------------------|
| 尺寸                       | 长 1.57 m, 直径 0.19 m | 工作水深 | 3.05—12.19 m (最大<br>91.44 m) |

|      |                 |        |                             |
|------|-----------------|--------|-----------------------------|
| 重量   | 42.64kg(最大)     | 电源     | 1kW 锂离子电池                   |
| 浮力   | 可调控 0~45%       | 发射平台   | 各种小型舰艇                      |
| 推进方式 | 电动主机/螺旋桨推进      | 频率(声学) | 900kHz 声呐<br>1200kHz 多普勒计程仪 |
| 数据通信 | 无线电—232/USB/以太网 |        |                             |

#### 性能数据:

|             |                 |
|-------------|-----------------|
| 定位精确度       | 14.94 m         |
| 探测/识别水雷的正确率 | 0.8 (A1 海区的沉底雷) |
| 探测/识别水雷的错误率 | 0.2 (A1 海区的沉底雷) |
| 可靠性         | 0.8             |
| 互用性         | 100%满足指定的标准     |

}

{

战场感知是战区无人系统的一个任务领域，无人系统的战场应用需经历一个分派、处理、利用和分发（TPED）过程，即将传感器探测数据转换成该环境下一种共享、可读性数据。其适用范围包括从执行空中和城市侦察任务，如空军的“捕食者”（Predator）、“死神”（Reaper）、“全球鹰”（Global Hawk），陆军的 PackBot 和 Talon 无人机，到未来作战任务，如远征军跑道评估、核侦察和特种部队海滩侦察作业。未来，无人技术将使执行任务的持久力从小时延长到天、周，从而使无人系统可进行更为持久、稳固的侦察和监视任务。由于无人系统将进一步向完全自主式方向发展，机载传感器必然要提供系统对自身环境的有机感知，故无论其意欲执行何种主要任务，均可附带对战场感知发挥作用。未来，完全自主的无人系统实质上能够同步执行战场感知的任务。战场感知任务领域实际是一种出让自身部分性能，从而支持跨多个战区或一个战区内，互相协作地执行任务和使命的工作方式。

}

entities.csv

| 实体 ID  | 实体名称               | 实体标签      |
|--------|--------------------|-----------|
| 000117 | MQ-9 “收割者” 无人机     | 型号无人机     |
| 000228 | ISR 任务             | 使命任务      |
| 004339 | AGM-114 “地狱火” 空地导弹 | 任务载荷-武器系统 |
| ...    | ...                | ...       |

entities\_attr.csv

| 实体 ID  | 实体属性 | 实体属性值 |
|--------|------|-------|
| 000117 | 翼展   | XX    |
| 000117 | 重量   | XX    |
| ...    | ...  | ...   |

relationships.csv

| 头实体 ID | 尾实体 ID | 关系 XX_1 | 关系 XX_2 | 关系 XX_3 |
|--------|--------|---------|---------|---------|
| 000117 | 000228 |         |         |         |
| 000117 | 004339 |         |         |         |
| ...    | ...    | ...     |         |         |

(关系 XX\_1、关系 XX\_2、关系 XX\_3，用于参赛队描述关系，可存放关系属性、关系标签等，由参赛队自行设计，也可以空着)

任务想定：

{

无人系统的 ISR 载荷有哪些？有何相似性？

}

## 2、任务要求

2.1 所构建知识图谱的实体，必须来自任务提供的原始数据，不允许使用原始数据之外的任何数据或外部知识库扩充实体规模。

2.2 对图谱构建过程中涉及的技术、方法、模型、训练数据等不作限制，但要求在最终提交的图谱构建技术报告中，清晰描述图谱构建过程与技术细节，以及数据来源（文件或地址），保证数据可获取、过程可复现，维护评测公平。

2.3 本次任务分图谱构建与图谱任务两个阶段进行，由于本次提供的外军无人系统数据源十分宝贵，为避免“无故套取数据”行为的再次发生，组织方将在图谱构建阶段，分两批发布原始数据，首批原始数据发布时间为 CCKS 评测任务数据发布时间，其后各参赛队根据该数据构建知识图谱，并将图谱文件提交[红山开源平台](#)进行评分；第二批原始数据，仅提供给有正常评分成绩的参赛队，这些参赛队再根据全部原

始数据，构建并向平台提交图谱文件。

2.4 “图谱构建阶段”提交的图谱文件，既是该阶段评分排名的依据，也是“图谱任务阶段”对所提交“任务分析报告”进行验证的依据，不允许根据“任务想定”修改图谱，再进行分析并提交“任务分析报告”。

### 3、数据描述

#### 3.1 原始数据

XX 原始数据.docx

#### 3.2 生成图谱数据

从原始数据文件中构建知识图谱，生成知识图谱存储在三份文件中：

entities.csv

relationships.csv

entities\_attr.csv

由于平台限制，只能上传一个 csv 文件，要求将 3 个待上传 csv 文件合并为一个 csv 文件，文件与文件之间用“###”作为行间隔隔开。

| 实体ID   | 实体名称             | 实体标签      |
|--------|------------------|-----------|
| 000117 | MQ-9“收割者”无人机     | 型号无人机     |
| 000228 | ISR任务            | 使命任务      |
| 004339 | AGM-114“地狱火”空地导弹 | 任务载荷-武器系统 |
| ....   |                  |           |
| ###    |                  |           |
| 头实体ID  | 尾实体ID            | 关系属性      |
| 000117 | 000228           | XX        |
| ....   |                  |           |
| ###    |                  |           |
| 实体ID   | 实体属性             | 实体属性值     |
| 000117 | 翼展               | XX        |
| 000117 | 重量               | XX        |
| ....   |                  |           |

如上图所示共有两行“###”，将上传的三个文件分隔为三部分。同一个文件列数必须相同，不同文件

间的列数不作要求。如上图中第一个 csv 文件有 3 列，第二个文件有 4 列（或 5 列）。

3 个文件顺序依次为：

entities.csv

relationships.csv

entities\_attr.csv

## 4、指标

本评测在构建阶段采用定量指标对图谱规模进行评分，指标具体定义如下：

遍历实体与关系列表，获取存在关系的实体的总数量  $N$ 。

遍历三份提交文件，获取全部三元组总数  $R$ 。

以  $ME = N + R$ ，为评测指标。

## 5、任务提交及评分

本次评测任务，分构建与任务两个阶段开展。

**第一阶段**为图谱构建阶段，时间为评测任务开始至“**任务想定**”发布之前，允许参赛队伍每天至多向平台提交 5 次结果，格式与任务描述中的示例输出相同，使用第 4 节所述的评价指标作为排名依据，排名每天更新一次。

**第二阶段**为图谱任务阶段，时间为“**任务想定**”发布至结果提交，各参赛队需提交以下材料以供审查：

(1) 图谱构建技术报告（构建方法的详细描述文档）

文档需包含图谱整体构建流程、各环节采用的具体方法及参数设置，若使用了额外资源，需做出说明并提供资源文件或地址。

(2) 任务分析报告

包括对“**任务想定**”的解析过程（给知识图谱提供的输入），得到的知识图谱输出内容及拓扑结构，对输出的分析结论。

**注意：**以参赛队在图谱构建阶段最后一次提交的图谱文件为最终图谱文件，任务分析报告必须来自最终图谱文件，不得根据“**任务想定**”修改图谱结构再进行分析，组织者将做专门审查。

结果提交方式（暂定，最终确定结果提交方式请关注公告通知）：将以上文件在结果提交截止日期前，发送至邮箱 lrp\_ph@163.com，邮件的标题为：“CCKS2022 图谱构建技术评测\_参赛队名称\_成果提交”。

## **6、奖励设置**

任务共提供 3.5 万元奖金给 4 支获奖团队，奖金设置如下：

第一名：12,000 元

第二名：10,000 元

第三名：8,000 元

创新奖：5,000 元

奖金及获奖证书由 CCKS2022 发放。

## **7、报名方式：**

鉴于本次评测的数据价值，同时也希望参赛队能切实参与评测，任务设置“报名-审批”环节。参赛队需首先在红山开源平台完成注册，并在“参赛报名”处填写以下信息：

- (1) 团队成员的姓名、联系电话、邮箱地址、所属单位；
- (2) 队长所在单位出具的证明（格式例如“我单位 XX，参加 XX 评测任务，特此证明”，加盖单位公章，公章不限单位级别），拍照作为上传文件，文件命名为：“CCKS2022 图谱构建技术评测\_参赛队名称\_报名”。

注：各参赛队提交报名信息后，将处于待审核状态，需经审核通过后才能视为报名成功。

## **8、时间安排**

时间安排初定如下，后续如果有调整，以邮件发布信息为准。

- 评测任务发布：4 月 25 日
- 数据发布：5 月 4 日
- 报名时间：5 月 4 日—6 月 25 日
- 任务想定发布：7 月 25 日
- 结果提交：7 月 31 日
- 评测论文提交：8 月 12 日
- CCKS 会议日期(评测报告及颁奖)：8 月 25 日—28 日

## **9、组织者信息**

任务组织者：

张 静，军事科学院系统工程研究院

任务联系人邮箱：(评测相关问题可通过该邮箱联系)

1rp\_ph@163.com