**CCKS 2023 技术评测任务书**

# 外军无人系统知识图谱推理问答评测任务

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，知识图谱是公认的智能化信息基础。组织方本着构建精准、可用的军事垂直领域权威知识图谱目标，自2020年开始，连续3年依托CCKS平台，组织无人系统知识图谱构建任务，着重提供权威、系统、精准的原始数据以及部分领域知识指导，目前，图谱知识库已初具规模。

中文知识图谱自然语言问答，简称CKBQA（Chinese Knowledge Base Question Answering），是连接具体用户和知识图谱之间的“人机接口”，是知识图谱服务能力发挥的重要支撑。年初以来，大语言模型在问答领域产生了颠覆性影响，但在机理上仍是基于数据的统计模型，与军事领域强调的“精准性”与“责任界面”要求存在一定差距。知识图谱背后的精准专家知识与可解释性，仍然是当前一段时间严肃数据分析、实际辅助决策等军事应用的主要支撑。为此，组织方开放前期通过评测任务构建的部分外军无人系统知识库，同步提供带有军事语言特点与无人系统关切点的问答数据集，通过组织技术评测，提升外军无人系统CKBQA的精准性，支撑知识图谱系统尽快上线服务用户，投入实践、发挥实用。

**1、任务定义**

1.1 输入

输入文件包含若干行自然语言问句。

1.2 输出

输出文件每一行对应一个问题的答案列表，列表内元素以\t 分隔。特殊地，若问题为布尔型问题，则输出 True 或 False 作为答案；若问题为计数型问题，则输出不带单位的整数作为答案。

1.3 输入样例

问题1：激光技术在军用无人系统领域的作用有哪些？

问题2：使用激光技术的军用无人机中，升限最高的是哪一型？

问题3：MQ-4C是 2013年首飞吗？

1.4输出样例

输出1：“激光制导\t激光通信\t激光测距”

输出2：“MQ-9”

输出3：“True”

**2、数据集描述**

2.1 数据来源

本次测评的问答数据采用专家介入与模板生成结合的方式构建，标注过程不依赖于特定模板，任务组织方邀请领域专家对语料风格、问题模式进行指导。问答数据集中的问题不局限于特定领域，且既包含简单问题（实体属性、关系的单一查询），也包含复杂问题（显示约束、隐式约束、比较、布尔、多跳等）。

2.2 训练集&验证集

在训练数据发布阶段，任务组织方会发布4000条标注好的数据（包括问题/SPARQL/答案）作为训练集。同时发布1000条左右不含标注结果的数据作为验证集，选手可以将自己生成的验证集答案提交，比赛系统会对答案进行评测，给出得分并进行排榜。

在测试数据发布阶段，任务组织方会发布验证集的标注结果（包括问题/SPARQL/答案），同时发布1000条左右不含标注结果的数据，作为最终测试数据。

训练文件会提供问题对应的Cypher查询。原则上不要求生成Cypher查询，参赛选手可以选择自己的方式，只需要给出问题的最终答案（答案需要来自给定知识库）。当问题所涉及的事实与知识库不一致时，以知识库中数据为准。

2.3 知识库相关文件说明

本任务使用的知识库主要由CCKS2021、CCKS2022两届外军无人系统知识图谱构建评测任务形成的知识图谱融合得到。知识库及说明，将发布在红山开源平台（https://www.osredm.com/competition/zstp2023/）。

**3、评价指标**

本任务的评价指标包括宏观准确率(Macro Precision)，宏观召回率(Macro Recall)，Averaged F1值。最终排名以Averaged F1值为基准。设$Q$为问题集合，$A\_{i}$为选手对第$i$个问题给出的答案集合，$G\_{i}$为第$i$个问题的标准答案集合，相关计算公式如下：

$Macro Precision=\frac{1}{|Q|}\sum\_{i=1}^{|Q|}P\_{i}$, $P\_{i}=\frac{|A\_{i}\bigcap\_{}^{}G\_{i}|}{|A\_{i}|}$

$Macro Recall=\frac{1}{|Q|}\sum\_{i=1}^{|Q|}R\_{i}$, $R\_{i}=\frac{|A\_{i}\bigcap\_{}^{}G\_{i}|}{|G\_{i}|}$

$$Averaged F1=\frac{1}{|Q|}\sum\_{i=1}^{|Q|}\frac{2P\_{i}R\_{i}}{P\_{i}+R\_{i}}$$

注：经过形式化生成方法，当前版本的数据不会出现一个问题对应多个答案的情况。

**4、任务提交及评分**

4.1 评测阶段

本次评测采取刷榜方式，分两个阶段开展。

第一阶段（训练验证阶段）：训练集与验证集发布后，允许参赛队伍多次向平台提交结果。参赛者将所有预测结果写入一个结果文件中上传给系统（每天提交不超过3次），如果新提交结果好于之前提交结果，排行榜中的成绩将自动进行更新覆盖。

第二阶段（测试排榜阶段）：测试集发布后，验证集提交通道将会关闭，其余事项与第一阶段相同。

4.2 最终提交材料

参赛队最终提交如下材料：

a) 问答任务测试集结果文件，用result.json命名（UTF-8格式），key 值应为问题 id，

value 值对应问题答案。

b) 相关代码及说明

c) 方法描述文档（非评测论文，评测论文撰写要求见CCKS 2023官网）

代码及其文档需打包成一个文件（rar、zip等均可），要求提交所有的程序代码及相关的配置说明，**程序应当可以运行且所得结果与result.json相符**。如果方法使用了额外资源，要求说明并提供资源文件或地址。

以上三个文件需在任务提交截止日期前发送至邮箱lrp\_ph@163.com,邮件的标题为：“CCKS2023军事语言图谱问答评测\_参赛队名称”。

4.3 依托平台

本次评测的任务提交，将依托红山开源平台（https://www.osredm.com/competition/zstp2023/）开展。

**6、任务要求**

a)在训练验证阶段未提交结果参与排榜的队伍，不允许继续参加测试阶段任务；

b)允许使用开源代码或工具，但不允许使用任何未公开发布或需要授权的代码或工具；

c)除主办方提供的数据集外，允许参赛队使用外部数据，但该数据必须是公开可获取的且不能用额外的标注数据，并在提交最终结果时一并提交外部数据及情况说明（如数据过大，需提供下载地址）；

d)不允许基于测试集或KB用规则Case By Case解决问题，一旦发现，任务组织方有权取消该队参赛资格；

e)参赛队最终需要提交可运行的代码和方法描述文档，若在排行榜上的结果无法复现，任务组织方将取消该队参赛资格。

**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（https://www.osredm.com/competition/zstp2023），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），拍照上传，文件命名为“CCKS2023军事语言图谱问答评测\_参赛队名称\_报名”。

参赛队通过审批环节后，可在红山开源平台完成数据集与知识库下载、提交计算结果、参与打分排榜。

**8、奖励设置**

本任务设置一、二、三等奖和技术创新奖，均获得中国中文信息学会颁发的获奖证书，并有机会录用EI检索的评测论文。

**9、时间安排**

时间安排初定如下，后续如果有调整，以CCKS发布信息为准。

* 评测任务发布：4月10日
* 报名时间：5月4日—7月14日
* 训练及验证数据发布：5月14日
* 测试数据发布：7月15日
* 提交测试结果：7月21日
* 评测论文提交：8月11日
* CCKS会议日期(评测报告及颁奖)：8月24日—27日

**10、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：**

**（可通过该邮箱联系沟通）**

**lrp\_ph@163.com**