**CCKS 2020 技术评测任务书**

# 面向试验鉴定的命名实体识别任务

军事装备试验鉴定是指通过规范化的组织形式和试验活动，对被试对象进行全面考核并作出评价结论的国家最高检验行为，涵盖方法、技术、器件、武器系统、平台系统、体系、训练演习等领域，涉及面广、专业性强。近年来，自然语言理解和人工智能技术飞速发展，日趋成为推动大数据建设的重要力量。试验鉴定由于试验目标的不同、被试对象的特点、武器系统的特性，有着自身较为特殊的语言形式，自然语言处理技术服务于该领域大数据建设的效果如何、基线在哪里，是各级都非常关注的问题。为此，我们以命名实体识别任务为切入点，逐步开展这方面的工作，以期通过评测的形式，促进交流、掌握基线、提升技术水平，遴选优秀技术为相关项目建设提供参考，共同推动试验鉴定领域大数据建设。

**1、任务定义**

**输入**：

1.试验鉴定相关自然语言文本$集合：D=\left\{d\_{1},\cdots d\_{N}\right\}, d\_{i}=\left〈w\_{i1},\cdots w\_{in}\right〉$

2.预定义类别：$C=\{c\_{1},\cdots c\_{m}\}$

**输出**：

实体提及和所属类别对的集合：$\{\left〈m\_{1},c\_{m\_{1}}\right〉,\left〈m\_{2},c\_{m\_{2}}\right〉,\cdots \left〈m\_{p},c\_{m\_{p}}\right〉\}$

其中$m\_{i}=\left〈d\_{i},b\_{i},e\_{i}\right〉$是出现在$文档d\_{i}$中的试验鉴定实体提及（mention），$b\_{i}$和$e\_{i}$分别表示$m\_{i}$在$d\_{i}$中的起止位置，$c\_{m\_{i}}\in C$表示所属的预定义类别。要求实体提及之间不重叠，即$e\_{i}<b\_{i+1}$。

多次出现的实体，只标注一次。

示例：

**输入：**美国洛马公司在新墨西哥州白沙导弹靶场，完成“微型碰撞杀伤”拦截弹重新设计后的第二次飞行试验，进一步检验了拦截弹的敏捷性和气动性能，标志着其成熟度进一步提升。“微型碰撞杀伤”拦截弹采取直接碰撞杀伤技术，主要用于提高美国陆军应对火箭弹威胁的能力。

**输出：**

"entities": [

{

"label\_type":"试验要素",

"overlap":0,

"start\_pos":22,

"end\_pos":32

},

{

"label\_type":"性能指标",

"overlap":0,

"start\_pos":57,

"end\_pos":59

},

{

"label\_type":"性能指标",

"overlap":0,

"start\_pos":61,

"end\_pos":64

},

{

"label\_type":"系统组成",

"overlap":0,

"start\_pos":92,

"end\_pos":99

},

{

"label\_type":"任务场景",

"overlap":0,

"start\_pos":107,

"end\_pos":110

},

{

"label\_type":"任务场景",

"overlap":0,

"start\_pos":113,

"end\_pos":117

},

]

**2、预定义类别**

预定义类别定义如下：

**1) 试验要素**：试验鉴定工作的对象，如列为考核目标的武器装备（系统级）、技术、战术、人员、对象之间的能力等；支持完成试验鉴定所需的条件，如陪试品、参试装备、测试、测量、靶标、仿真等；装备的基本情况等。

例如：RS-24弹道导弹、SPY-1D相控阵雷达、紫菀防空导弹（Aster）、F-35“闪电”II型联合攻击战斗机、“阿利·伯克”级Flight IIA型驱逐舰“约翰芬”号、协同通信与指挥、连续波测量雷达、电影经纬仪、无人机靶标等。

**2) 性能指标**：试验要素在技术、使用等性能方面的定性、定量描述，如重量、射程、可靠性等。

例如：测量精度、圆概率偏差、失效距离、准备时间、反激光毁伤、发射方式等。

**3) 系统组成**：被试对象的组成部分，如子系统、部件、采用的技术等。

例如：动能杀伤飞行器（KKV）、中波红外导引头、助推器、整流罩、箔条红外混合诱饵弹、碰撞杀伤技术、柔性摆动喷管技术、端羟基聚丁二烯、等。

**5) 任务场景**：试验要素在发挥其实际效用和价值中涉及的信息，如人员、对抗目标、体系能力等。

例如：法国海军、导弹预警、恐怖袭击、迫击炮威胁、排级作战等。

**3、数据集**

**3.1 数据标注说明**

由于该领域特殊性及保密性，开放语料非常匮乏。军事科学院系统工程研究院组织对外军装备试验鉴定资料进行人工标注，通过独立标注、交叉检查、专家复核、修改迭代的方式，保证标注一致性。该数据集仅限CCKS竞赛评测用。

**3.2 数据示例**

{

"originalText":"extron公司于当日宣布,成功试射了其研制的精确制导滑翔武器——G-CLAW,并命中移动目标.此次试验于2016年10月在尤马靶场进行.试验中,导弹从塞纳斯大篷车系列飞机上发射,通过将GPS半自动惯性导航系统转换为半主动激光传感器末端制导,以1米以内的圆概率误差分别命中静止和动态目标.G-CLAW 精确制导滑翔武器系统融合了3种引信模式,能够针对不同目标选择更好的模式.该武器采用模块化设计,以支持其与制导\传感和弹头效应等方面的新兴技术快速融合.",

"entities":[

{"label\_type":"试验要素","overlap":0,"start\_pos":24,"end\_pos":39},

{"label\_type":"试验要素","overlap":0,"start\_pos":44,"end\_pos":47},

{"label\_type":"试验要素","overlap":0,"start\_pos":77,"end\_pos":86},

{"label\_type":"性能指标","overlap":0,"start\_pos":128,"end\_pos":132},

{"label\_type":"性能指标","overlap":0,"start\_pos":137,"end\_pos":138},

{"label\_type":"性能指标","overlap":0,"start\_pos":140,"end\_pos":143},

{"label\_type":"系统组成","overlap":0,"start\_pos":94,"end\_pos":105},

{"label\_type":"系统组成","overlap":0,"start\_pos":109,"end\_pos":116},

{"label\_type":"系统组成","overlap":0,"start\_pos":193,"end\_pos":197},

{"label\_type":"任务场景","overlap":0,"start\_pos":167,"end\_pos":170},

{"label\_type":"任务场景","overlap":0,"start\_pos":176,"end\_pos":179},

{"label\_type":"任务场景","overlap":0,"start\_pos":204,"end\_pos":213},

{"label\_type":"任务场景","overlap":0,"start\_pos":218,"end\_pos":221}

]

}

**3.3数据集描述**

本次评测的训练数据有：

1. 400条标注数据
2. 4个类别的3398个实体词词表

标注数据集统计如下表（最终结果以发布数据为准）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 　 | 文本 | 试验要素 | 性能指标 | 系统组成 | 任务场景 | 总数 |
| 训练集 | 400 | 1188 | 854 | 694 | 662 | 3398 |

**4、评价指标**

本次任务，采用F1-Measure作为评测指标，具体定义如下：

试验鉴定相关自然语言文本$集合：D=\left\{d\_{1},\cdots d\_{N}\right\}, d\_{i}=\left〈w\_{i1},\cdots w\_{in}\right〉$

预定义类别：$C=\{c\_{1},c\_{2},c\_{3},c\_{4}\}$

对于任意一个预定义类别*c*i，设集合*D*中包含*c*i的实体数量为AP，对*c*i进行识别，正确识别结果数量为TP，错误识别结果数量为FP。

精确率P定义为： P = TP/(TP+FP)

召回率R定义为： R = TP/AP

F1值定义为：F1=2\*P\*R/(P+R)

**5、奖励设置**

除CCKS组委会设置的1,5000元奖金外，军事科学院系统工程研究院额外提供1,5000元奖金给获奖团队，奖金设置如下：

第一名：12,000元

第二名：7,000元

第三名：4,000元

技术创新奖：7,000元

**6、任务提交指南**

本次评测任务依托Biendata平台展开,将采取刷榜的方式，验证集发布后，允许参赛队伍每天至多向平台提交一次结果，格式与任务描述中的示例输出相同，取四个预定义类各自F1值的算数平均值，作为排名依据，排名每小时更新一次。参赛队伍可在测试集发布之前随时上传验证集的计算结果，平台管理系统会及时更新各队伍的最新排名情况，验证集榜不计入比赛最终成绩。

测试集为最终评估比赛成绩的有效评测集，待测试集发布后，允许参赛队伍每天至多提交一次，总共至多提交五次评测结果，取五次评测结果中成绩最高的一次，作为评比成绩。

评比成绩排名进入前八名的参赛队伍，提交以下材料以供最终审查：

(1) 测试集结果文件，用“参赛队名称\_TE\_result.txt”命名（UTF-8格式）

(2) 相关代码及文档

(3) 方法描述文档（非评测论文，评测论文撰写要求见CCKS 2020官网）

代码及其文档需打包成一个文件（tar，zip，gzip，rar等均可），作为邮件附件传送，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.txt相符。方法描述文档需包含算法描述及参数设置。**如果方法使用了额外资源，需做出说明并提供资源文件或地址。**

将以上三个文件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：“CCKS2020-TE\_NER-参赛队名称”。

**7、时间安排**

时间安排初定如下，后续如果有调整，将在讨论组中更新说明。

* 评测任务发布：3月20日
* 报名时间：3月20日—7月10日
* 训练及验证数据发布：3月30日
* 测试数据发布：7月11日
* 提交测试结果：7月11日-7月15日
* 评测论文提交：8月5日
* CCKS会议日期(评测报告及颁奖)：8月15日—18日

**8、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：**

**栾瑞鹏，lrp\_ph@163.com**

**孙 晓，psunxiao@163.com**