CCKS2024 大模型零 样 本 知 识 抽 取 评 测

# 任务简介

随着深度学习和预训练模型技术的飞速发展，零样本知识抽取正吸引着越来越多的关注。在零样本知识抽取任务中，我们探究如何利用大型预训练语言模型如ChatGPT、Mistral、LLaMA、ChatGLM、文心一言、通义等，在没有任何针对特定任务标注数据的情况下，提取文本中的结构化信息。这些先进的语言模型内嵌了丰富的世界知识，能够通过其庞大的参数化网络对文本进行理解和分析，体现出它们在知识获取和任务泛化方面的优势。

零样本知识抽取聚焦于设计策略，使得大型预训练模型能够直接对新领域或不熟悉的任务进行操作，无需依赖手动标注的训练数据集。应用场景涵盖从非结构化文本中识别命名实体、提取关键信息、确定实体之间的关系、甚至是抽取更复杂的事件和情感等信息。

本次任务的目标是评估并进一步促进大模型在零样本场景下的知识抽取能力，在没有专门针对性标注数据的支持下，挖掘模型处理不同类型和来源的文本所蕴含的潜力。这不但有助于减少知识抽取任务中的人力标注成本，同时也推动了对预训练模型语义理解和泛化能力的深入理解与探索。

为激发研究者提出创新的零样本知识抽取方法，同时，在没有标注数据的条件下，为实际应用场景提供有效的知识抽取技术方案，浙江大学和蚂蚁集团在CCKS2024大会组织本次评测任务。本次评测将依托阿里巴巴天池平台（<https://tianchi.aliyun.com/>）展开。

# 赛程安排

本次大赛分为初赛和复赛两个阶段，具体安排和要求如下：

报名组队——————5月10日—7月15日

初赛阶段——————5月10日—7月15日

复赛阶段——————7月15日—8月15日

评测论文——————9月3日

评测颁奖——————9月19日-22日

报名组队与实名认证（2024年5月10日—7月15日）

报名方式：5月10日阿里天池平台(<https://tianchi.aliyun.com/>)将开放本次比赛的组队报名、登录比赛官网，完成个人信息注册，即可报名参赛；选手可以单人参赛，也可以组队参赛。组队参赛的每个团队2-3人，每位选手只能加入一支队伍；

选手需确保报名信息准确有效，组委会有权取消不符合条件队伍的参赛资格及奖励；复赛前3名队伍将被要求提交代码进行复现审核，如果无法提交代码或结果相差较大将依次递补。

选手报名、组队变更等操作截止时间为7月15日23：59：59；各队伍（包括队长及全体队伍成员）需要在7月15日23：59：59前完成实名认证（认证入口：天池官网-右上角个人中心-认证-支付宝实名认证），未完成认证的参赛团队将无法进行后续的比赛；

大赛官方钉钉群请搜索Group Number：63840014837，或扫描以下二维码加入，最新通知将会第一时间在群内同步：

钉钉群二维码

初赛阶段（2024年5月10日-2022年7月15日，UTC+8）

初赛的几个关键时间点： 5月10号天池平台将开放竞赛数据集和系统测评。

选手报名成功后，参赛队伍通过天池平台下载数据，本地调试算法，在线提交结果。

初赛仅提供测试数据集，供参赛选手提交评测结果,参与排名。初赛阶段提交格式在具体任务章节有详细介绍。

初赛时间为2024年5月10日-2024年7月15日，系统每天提供2次评测机会，系统进行实时评测并返回成绩，排行榜每小时进行更新，按照评测指标从高到低排序。排行榜将选择参赛队伍在本阶段的历史最优成绩进行排名展示。

本次评测初赛仅供热身，不淘汰队伍。

复赛阶段（2024年7月15日—2024年8月15日，UTC+8）

复赛阶段会提供复赛测试数据集下载，供参赛选手提交评测结果，参与排名。复赛阶段提交规范和初赛阶段保持一致。

复赛时间为2024年7月15日-2024年8月15日。本阶段，系统每天提供5次实时评测，每小时更新排行榜，按照评测指标从高到低排序。排行榜将选择参赛队伍在本阶段的历史最优成绩进行排名展示。复赛提交截止时间8月15日17：59：59。  
 复赛结束后，大赛组织方将通知top3优胜队伍提交代码（提交要求见具体的子任务章节），赛题组织方将对优胜参赛队伍代码进行审核，要求模型能复现出最优提交成绩。对于未提交、复现未成功或审核不通过的队伍，将取消资格和比赛奖励，并通知递补选手。最终队伍排名按照复赛成绩核算，榜单将在官方完成代码审核后公布。

另：本次赛事若要求论文环节，则将以CCKS 2024的通知要求为准。

# 规则

1. 所有参赛选手都必须在天池平台管理系统中注册；

2. 参赛选手需确保注册时提交信息准确有效，所有的比赛资格及奖金支付均以提交信息为准；

3. 参赛选手在管理系统中组队，参赛队伍成员数量不得超过3个，报名截止日期之后不允许更改队员名单；

4. 每支队伍需指定一名队长，队伍名称不超过15个字符，队伍名的设定不得违反中国法律法规或公序良俗词汇，否则组织者有可能会解散队伍；

5. 每名选手只能参加一支队伍，一旦发现某选手以注册多个账号的方式参加多支队伍，将取消相关队伍的参赛资格；

6. 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；

7. 除主办方提供的数据集外，参赛选手允许使用公开可获得的预训练数据（如词向量、字向量等），允许使用公开的数据集（如InstructIE、IEPile等）对大模型进行预训练或微调来提升知识抽取能力（训练数据在审核代码阶段需提供给组织方进行审核！），评测严禁标注测试集或构造测试集相似数据进行训练。

8. 参赛队伍可在参赛期间随时上传验证集的预测结果，初赛阶段每天2次机会，复赛阶段每天5次机会。管理系统会定时更新各队伍的最新排名情况。

# 奖励设置

本次任务总奖金池为30000元，奖金设置如下：

第一名：人民币12000元，1名

第二名：人民币8000元，1名

第三名：人民币5000元，2名

# 任务组织者和联系人

**任务组织者：**

张宁豫，浙江大学

袁琳，蚂蚁集团

桂鸿浩，浙江大学

孙梦姝，蚂蚁集团

罗玉洁，浙江大学

徐军，蚂蚁集团

方继展，浙江大学

渠源,蚂蚁集团

薛逸达，浙江大学

梁磊，蚂蚁集团

**学术指导组**

陈华钧，浙江大学

周俊，蚂蚁集团

漆桂林，东南大学

王昊奋，同济大学

**任务联系人：**

桂鸿浩：[guihonghao@zju.edu.cn](mailto:guihonghao@zju.edu.cn)

张宁豫：[zhangningyu@zju.edu.cn](mailto:zhangningyu@zju.edu.cn)

# 支持社区

中文开放知识图谱OpenKG

# 7.任务描述

### 在零样本知识抽取任务中，参与者面对的挑战是从给定的文本中识别和提取指定类型的信息，而无需依赖事先标注的训练数据。此任务要求模型能够理解和遵循抽取指令—一个明确的命令，指导模型找到并格式化所需信息。

#### 输入

### 任务输入包括两部分内容：一是需要处理的文本（input），二是定义了输出结构和内容要求的抽取指令（instruction）。

#### 输出

### 任务输出是一个结构化的结果，根据抽取指令中提出的格式要求生成。输出需要准确反映从输入文本中抽取的信息。任务目标

## 8.任务特点

任务中的抽取指令涵盖了高度多样化的Schema，这一特点使得它与那些通常只需抽取少数几种固定类型信息的任务形成了鲜明对比。这种多样性要求模型具备灵活适应和理解各类抽取需求的能力。

每一项抽取任务均提出了明确的输出格式要求。参与者需要严格按照这些格式要求来组织和呈现抽取结果，确保输出的信息结构化程度和准确性。

抽取指令中不仅对输出格式提出了要求，还明确定义了需要抽取的Schema。参与者必须准确理解这些定义，并据此解读指令进行知识抽取。

这三点特性共同构成了本任务的独特挑战，要求参与者在执行知识抽取时，不但要精确理解和遵循抽取指令的要求，同时也需要具备高度的适应性和灵活性，以应对各种不同的抽取需求和场景。

### 数据集介绍

### 初赛测试集

初赛使用。

|  |  |
| --- | --- |
| 任务 | 零样本知识抽取 |
| 数据集 | 百科人物 |
| 测试集大小 | 1000 |

**test.json**

测试集文件test.json包含若干样本，每个样本拥有id、instruction和label三个字段。其中，instruction字段是模型的输入指令，而label字段是该输入的标注答案。示例数据如下：

{

"id": 28542,

"instruction": "{\"instruction\": \"你是一个图谱实体知识结构化专家。根据输入实体类型(entity type)的schema描述，从文本中抽取出相应的实体实例和其属性信息，不存在的属性不输出, 属性存在多值就返回列表，并输出为可解析的json格式。\", \"schema\": [{\"entity\_type\": \"足球运动员\", \"attributes\": {\"场上位置\": \"地位或角色。\", \"惯用脚\": \"靠习惯使用的脚来完成任务或活动的动作。\", \"出生地\": \"人物出生的具体地点。\"}}], \"input\": \"霍尔岑拜因，194年3月9日出生于德国法兰克福，场上司职前锋，曾效力于法兰克福、克鲁塞罗足球俱乐部。\"}",

"label": "{\"足球运动员\": {\"霍尔岑拜因\": {\"场上位置\": \"前锋\", \"出生地\": \"德国法兰克福\"}}}"

}

### 复赛测试集

待发布。

### 评测指标

本次任务采用F1值来评估模型进行零样本知识抽取的效果。

### 基线模型

OneKE是由蚂蚁集团和浙江大学联合研发的大模型知识抽取框架，具备中英文双语、多领域多任务的泛化知识抽取能力，并提供了完善的工具链支持。OneKE已经以开源形式贡献给OpenKG开放知识图谱社区。

OneKE模型介绍页面：<https://github.com/zjunlp/DeepKE/blob/main/example/llm/OneKE.md>

其他基线参考代码：<https://github.com/zjunlp/DeepKE/tree/main/example/llm/InstructKGC>

OneKE模型在测试集上的评估结果如下所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | P | R | F1 |
| baseline | 73.36 | 52.68 | 61.32 |

本次评测用户可使用任意公开的抽取数据集对大模型进行优化，如：

InstructIE：<https://wisemodel.cn/datasets/zjunlp/InstructIE> (训练集存在一定程度噪音)

IEPle: <https://github.com/zjunlp/IEPile> （由公开数据集合并处理得到）