**CCKS 2020 技术评测任务书**

# 新冠知识图谱构建与问答评测

## 评测背景

随着互联网软硬件相关技术的飞速发展，人们逐渐从信息时代进入智能时代。知识图谱作为承载底层海量知识并支持上层智能应用的重要载体，在智能时代中扮演了极其重要的角色。而受限于非结构化文本和结构化知识之间的巨大差异，自动化构造知识图谱以及利用知识图谱支持上层应用仍存在诸多挑战。因此对知识图谱的构建以及其核心应用基于知识图谱的自然语言问答展开研究是十分有必要的。

知识图谱又可以划分成开放领域和特定领域。相对而言，开放领域知识图谱的特点是“广而浅”，即覆盖实体的范围广，但可能在某方面缺少深层次或不常见的知识；而特定领域知识图谱的特点是“专而深”，即只覆盖特定领域的实体信息，但会涵盖一些通常只有对应领域专业人士使用的术语、概念以及对应的知识。随着LOD、OpenKG等知识共享项目的展开，将包括开放领域和特定领域的若干知识库融合或链接在一起从而形成既全面又深入的“超级”知识图谱逐渐成为可能。

时值2020年新型冠状病毒疫情爆发，在OpenKG总体组织和协调下，部分相关企业院校使用自动化的技术，以新型冠状病毒为核心构建了包括新冠百科、健康、防控等多个高质量的知识图谱（<http://openkg.cn/group/coronavirus>），并于此针对知识图谱构建的关键技术及其核心应用提出四个评测子任务。通常构建知识图谱需要在实体识别完成后，为每一个实体分配预定义的类型，这即是我们提出的第一个任务“**实体类型推断**”。而人工预定义的实体类型覆盖程度有限且不易更新，当涉及新的领域时，实体类别体系可能需要重新定义。通过在网络中动态的获得实体的概念类别，并自动化识别类别之间的上下位关系不但可以解决人工预定义的缺陷，还可以使知识图谱更加立体丰满，有助于上层应用。这即是我们提出的第二个任务“**概念的上下位关系预测**”。而为了使知识图谱中的各实体相连起来，需要判断哪些实体之间存在关系，以及实体通过特定关系可以链接到哪些实体，这即是我们提出的第三个任务“**链接预测**”。最后，当知识图谱构建完成后，使用自然语言对其进行查询可以极大地方便用户获取想要的知识，同时诸如智能音箱等应用也需要允许用户使用自然语言与知识图谱进行交互，这即是我们提出的第四个任务“**知识图谱的自然语言问答**”。

以下分别对本评测四个任务进行详细介绍。

## 任务描述

### 子任务一：新冠百科知识图谱类型推断

类型信息在知识库中具有非常高的价值，实体类型推断的研究一直是领域的热点。然而，大量类型信息以非结构化文本形式呈现于网络页面中，文本处理难度大，抽取结果同时保证高准确度和覆盖率仍然是个极大的挑战。针对实体的通用类型推断，近年来已有若干解决方案，如使用统计机器学习方法及利用外部知识（通向其他数据源的链接或文本信息）等。

本评测任务围绕新冠百科知识图谱构建中的实体类型推断（Entity Type Inference）展开。评测从实体百科（包括百度百科、互动百科、维基百科、医学百科）页面出发，从给定的数据中推断相关实体的类型。评测本身不限制各参赛队伍使用的模型、算法和技术。希望各参赛队伍发挥聪明才智，构建各类无监督、弱监督、远程监督、半监督等系统，实现知识图谱的实体类型推断，共同促进知识图谱技术的进步。

#### 输入输出

* **输入**

entity.txt：需要进行类型预测的所有实体，其中包含相关实体和噪音实体。

entity\_pages\_1.xml，entity\_pages\_2.xml，entity\_pages\_3.xml，entity\_pages\_4.xml：分别来自百度百科，互动百科，中文维基百科，医学百科的实体页面内容。保证entity.txt中的任意实体至少被一个页面文件所涵盖。实体页面文件中包含类型推断可能会用到的名称、标签、简介等信息

type.txt：包含7个目标类型，分别是病毒、细菌、疾病、药物、医学专科、检查科目、症状。

* **输出**

entity\_type.txt：实体及预测的实体类型，每行一个“实体\t类型”对。注意不要包含类型不在给定的7种类型当中的实体。

* **输入样例**

entity.txt:

 烟草花叶病毒

 大肠杆菌

 艾滋病

 盐酸西普利嗪

内科

太阳

entity\_page\_1.xml(页面文件均与之类似，这里省略其他三个百科的实体页面)



type.txt：

病毒

细菌

疾病

药物

医学专科

检查科目

症状

* **输出样例**

 烟草花叶病毒 病毒

 大肠杆菌 细菌

 艾滋病 疾病

 盐酸西普利嗪 药物

内科 医学专科

* **说明**

实体“太阳”不在给定的7种目标类型当中，因此输出样例中不包含该实体。

#### 评价指标

本任务采用精确率（Precision, P）、召回率（Recall, R）、F1值（F1-measure, F1）来评估效果。设$A$为参赛队伍输出文件中的所有实体-类型对的集合，$G$为评测方标注文件中的所有实体-类型对的集合，相关计算公式如下：

$$P=\frac{|A\bigcap\_{}^{}G|}{|A|} R=\frac{|A\bigcap\_{}^{}G|}{|G|} F1=\frac{2PR}{P+R}$$

#### 数据集

本任务总共包含17.5W左右的实体页面。任务本身不限定方法类型，可以是无监督、半监督、有监督方法，因此不提供任务相关训练集，参赛选手如有需要，可以自行对训练数据进行标注。

测试集是主办方通过自动化实体类型推测和人工检验进行标注的。测试集上不允许参赛选手进行任何人工标注。

### 子任务二：新冠概念图谱的上下位关系预测

在如今的信息化时代，互联网中实体类别多样化，且粒度更细并具有层次，相对于类别有限的传统命名实体，人们开始将目光转向开放域实体挖掘。

新冠概念图谱即是一种自动构建概念体系（schema）的知识图谱，该图谱利用自动挖掘的手段从网络文本中采集了大量细粒度的上位概念词。即此概念图谱的is-a层次结构（是自动构建而成的，包含了细粒度的上下位层次结构。

传统的知识图谱对实体的概念类别体系定义非常有限，如：ACE-2007将实体分为7大类、45小类，Yosef（2013）将实体分为505类。上述将命名实体的类别进行人为的定义，其优点在于可以将命名实体识别中标明实体类别的过程看作分类问题，然后应用传统的模式分类方法解决该问题。然而，预先对类别进行定义也有其不可避免的缺陷：人工定义的类别覆盖程度有限且不易更新，当涉及新的领域时，实体类别体系可能需要重新定义。在如今的信息化时代，对于互联网中的海量实体很难由人工预先定义出一个完备的类别体系，基于此，可以由网络中动态的获得实体的概念类别，这种动态挖掘的方法有以下两个主要特点：

(1) 实体词的类别更多，且不限定。比如可能的类别包括药品、动物、植物、赛事、会议、菜肴等等，远远多于传统命名实体。而且随着社会的进步，一些新的类别可能出现，因此靠人工难以确定一种固定的完备的类别体系。

(2) 实体词的类别粒度更细，且有层次。比如传统命名实体中的机构名可以进一步细分为：学校名、公司名、政府部门名、新闻机构名等；学校名则还可以继续细分为高校名、中学名、小学名等。这些类别通过上下位关系连接，构成一种偏序结构。

事实上，实体和上位词之间以及上位词和上位词之间存在复杂的层次关系，如果能够自动的构建这种细粒度的上下位层次关系能够为众多的智能服务系统提供帮助。

#### 输入输出

* **输入**

entity.txt：实体列表。

concept.txt：概念（类型）列表。

* **输出**

entity\_concept.txt：实体-概念之间的类型关系。

concept\_concept.txt：概念-概念之间的上下位关系，前者是后者的子概念。

* **输入样例**

entity.txt:

 钟南山

新冠病毒

concept.txt：

病毒

细菌

疾病

药物

医学专科

检查科目

症状

…

* **输出样例**

entity\_concept.txt:

 钟南山 著名医学家

钟南山 教授

新冠病毒 病毒

concept\_concept.txt：

著名医学家 医学家

医学家 专家

专家 人物

教授 教师

教师 人物

病毒 微生物

微生物 生物

#### 评价指标

本任务采用精确率（Precision, P）、召回率（Recall, R）、F1值（F1-measure, F1）来评估效果。

对于Entity-Concept（实体-概念）类型关系，设$E$为测试集中的实体集合（注：测试集是公开的实体集的子集），$A\_{i}$为选手对$E$中第$i$个实体给出的类型集合，$G\_{i}$为第$i$个实体的正确类型集合，相关计算公式如下：

$Macro Precision(EC)=\frac{1}{|E|}\sum\_{i=1}^{|E|}P\_{i}$, $P\_{i}=\frac{|A\_{i}\bigcap\_{}^{}G\_{i}|}{|A\_{i}|}$

 $Macro Recall(EC)=\frac{1}{|E|}\sum\_{i=1}^{|E|}R\_{i}$, $R\_{i}=\frac{|A\_{i}\bigcap\_{}^{}G\_{i}|}{|G\_{i}|}$

$$Averaged F1(EC)=\frac{1}{|E|}\sum\_{i=1}^{|E|}\frac{2P\_{i}R\_{i}}{P\_{i}+R\_{i}}$$

对于Concept-Concept（概念-概念）上下位关系，设$A$为参赛队伍提交文件中的所有概念-概念对的集合，$G$为评测方标注文件中的所有概念-概念对的集合，相关计算公式如下：

$$P(CC)=\frac{|A\bigcap\_{}^{}G|}{|A|} R(CC)=\frac{|A\bigcap\_{}^{}G|}{|G|} F1(CC)=\frac{2PR}{P+R}$$

最终得分 $F1\left(overall\right)=\left(Averaged F1\left(EC\right)+F1\left(CC\right)\right) / 2$

#### 数据集

本任务共公开20000左右实体，1000左右概念。任务本身的设置是无监督的，因此不提供训练集。测试集是主办方通过自动化实体类型推测和人工检验进行标注的，采用公开数据集的子集作为测试集（500左右实体，500左右概念）。比赛期间不公开测试集。根据选手提交的输出和人工标注的测试集进行比对输出最后结果。

### 子任务三：新冠科研抗病毒药物图谱的链接预测

目前，由于新冠病毒的大范围传播，对人类健康和社会经济造成了重要的影响，因此抗病毒药物的研发变得尤为重要。病毒的核心是核酸分子（RNA或DNA）和蛋白质外壳，其寄生于宿主体内，通过一系列步骤进行复制增值。抗病毒药物可以靶向病毒复制的某个环节，发挥抗病毒作用。抗病毒药物研发中，研究者需要对各种抗病毒药物做大量生物实验，从而确定病毒抑制效果做，但是此步骤需花费大量时间。在大数据和人工智能时代，我们可以收集大量的临床和实验数据，利用AI算法对抗病毒药物和病毒的靶向作用、病毒蛋白和宿主蛋白的交互作用等进行预测。其中，抗病毒药物图谱是解决该问题的一种途径，通过构建抗病毒药物知识图谱，利用大规模图算法以及知识推理来预测抗病毒药物和病毒的靶向作用、病毒蛋白与宿主蛋白的交互作用。

抗病毒药物图谱是依据抗病毒药物、病毒、病毒相关蛋白和宿主及宿主蛋白间的关系来构建的知识图谱。抗病毒药物图谱可以为抗病毒药物的生物研究提供一定支持，用于辅助预测抗病毒药物对病毒的影响、病毒蛋白和人类蛋白的交互作用等问题。这类问题对抗病毒药物的研发和基于知识图谱的推理具有重大的实际意义和学术价值。

本任务基于抗病毒药物图谱进行关系预测，如药物和病毒的靶向作用、蛋白间的交互作用等。评测依据抗病毒药物图谱模式（Schema）和知识图谱的实体、实体属性、实体之间的关系，预测新的两个实体的关系。评测本身不限制各参赛队伍使用的模型、算法和技术。可以利用各种图算法模型，预训练等手段进行图谱上的关系预测，共同促进知识图谱技术的发展。

#### 输入输出

* **输入**

schema.json：知识图谱的图谱模式（Schema），即定义了知识图谱的实体类型（Entity）、实体属性名（Attribution key）和实体间的关系（Relationship）。

entities.json：实体列表，即病毒、药物、宿主蛋白、病毒蛋白等实体。

attrs.json：实体属性列表，即病毒的类型、药物的类型等等。

relationships.json：实体与实体之间的关系列表，即病毒-药物作用，病毒-病毒蛋白关系，病毒蛋白-宿主蛋白的关系等。

link\_prediction.json：待评测的头实体或者尾实体和关系组成的对。

* **输出**

result.txt：针对每个部分缺失的实体关系对所预测的对应top 10头或尾缺失实体队列集合（按可能性从高到低排序）。**选手需要从队列中删除训练集已有链接的实体数据，防止占用并浪费队列位置。**

* **输入样例**

schema.json：

{

 "entity\_type": ["virus", "drug", ……],

 "attrs": {

 "virus": {"name": "string", "class": "string", ……},

 "drug": {"name": "string", "indication": "string", "drug\_type": "string", ……},

 ……

 }，

 "relationships": [["drug", "effect", "virus"], ["virus", "produce", "virusProtein"], ["HostProtein", "interaction", "virusProtein"], ……],

}

entities.json：

{

 "virus": ["Human papillomavirus 9", "Human adenovirus 28", "Equine arteritis virus Bucyrus", ...],

}

attrs.json:

{

"attrs": [["Equine arteritis virus Bucyrus", "alias", " Equine arteritis virus (strain Bucyrus) "], ...]

}

relaionships.json

{

 "relationships": ["Equine arteritis virus Bucyrus", "produce", "RPOA"]

}

link\_prediction.json

{

 "relationships":

 [

 ("Equine arteritis virus Bucyrus", "produce", "?"),

 ("?", "produce", "RPOA"),

 ……

 ]

}

* **输出样例**

link\_prediction.json:

{

"results":

[

[A1, B1, C1, D1, …],

[A2, B2, C2, D2, …],

……

]

}

#### 知识图谱描述

如下图可视化展示，effect关系用于表示抗病毒药物对于病毒有一定效果，例如Abacavir和Lopinavir对Human immunodeficiency virus 1有抗病毒效果，通过本关系可查看抗病毒药物和对应病毒间的联系；



produce关系可用于挖掘病毒与其所表达蛋白间的express关系，并可进一步挖掘不同病毒与同一类型蛋白间的关系，进而发现两个病毒间的间接关系；再通过某病毒与某抗病毒药物间关系，又可发现某抗病毒药物和另一病毒的间接关系；可以帮助发现对于某一类型病毒蛋白可能有作用的更多药物。

Interaction和binding关系可以用于发现病毒蛋白与宿主蛋白间的相互作用，比如HIV1的NP（nucleoprotein）和Human的HYOU1 和间的binding关系；还有某些药物和HYOU1间的target关系，进而可以研究这些药物和HIV1的NP（nucleoprotein）间的作用。

通过belong to关系，可以明确哪些蛋白属于哪类宿主，这类宿主的哪些蛋白与病毒有相互作用关系等。在属性方面，我们又进一步从DrugBank挖掘了这些信息（商品名、化学式，分子重量，indication等），作为drug实体的属性，使得研究者可以更直观的了解某drug的一些重要信息。

#### 评价指标

本次评测任务借鉴MRR (Mean reciprocal rank，平均倒数排名) 来评估构建效果，我们考虑多种关系预测：病毒-药物关系预测、蛋白-蛋白交互预测，病毒-病毒蛋白交互预测等，这些共同构成全局的关系预测。相关的定义如下：

MRR (Mean reciprocal rank，平均倒数排名)：对于一个query，若第一个正确答案排在第n位，则MRR得分就是 $1/n$。

$$MMR=\frac{1}{|Q|}\sum\_{i=1}^{|Q|}\frac{1}{rank\_{i}}$$

其中，Q为样本query的集合（*link\_prediction.json*），|Q|表示query的个数。

$$\frac{1}{rank\_{i}}=\left\{\begin{array}{c}\frac{1}{i} (目标实体在第 i个结果中命中)\\0 (目标实体在所有结果中未命中)\end{array}\right.$$

每轮最多一次命中

#### 数据集

**本任务不允许使用公开的外部知识库。**

本任务目前总共包含8000左右实体，50000左右的三元组。

数据集来源于正规网站的结构化数据，训练集与测试集的比例约15 : 1 。

### 子任务四：新冠知识图谱问答评测

本任务属于中文知识图谱自然语言问答任务，简称CKBQA （Chinese Knowledge Base Question Answering）。即输入一句中文问题，问答系统从给定知识库中选择若干实体或属性值作为该问题的答案。问题均为客观事实型，不包含主观因素。理解并回答问题的过程中可能需要进行实体识别、关系抽取、语义解析等子任务。这些任务的训练可以使用额外的资源，但是最终的答案必须来自给定的知识库。

知识图谱问答在当前互联网信息爆炸、人工智能盛行的时代是十分有战略价值和研究意义的。一方面，传统搜索引擎是以网页资源为核心，依据关键词索引、文本匹配等方式进行检索并返回给用户相关网页链接，而用户很多时候需要的只是对一个具体问题的特定解答。另一方面，诸如智能音箱、智能问诊等新一代产品应用通常也需要依赖于特定的知识图谱响应用户的自然语言请求。例如智能问诊应用可以依据医药健康领域的知识库对患者的情况进行初步诊断。

本次知识图谱问答任务是在CCKS上举办的第三届。我们结合今年的时事，以发布在OpenKG上的新冠开放知识图谱为核心，以新型冠状病毒为切入点悉心构造了一定规模的针对健康、医药、疾病防控等特定主旨的问答数据。同时依然保留去年开放领域的问答数据以供参赛队伍对模型进行训练。我们期望参赛选手的问答系统既能处理各种百科类的浅层问题，也能处理具备一定领域知识（如流行疾病等）的较深层问题。在后续的任务中，我们可能会针对其他领域加入更多的问答数据。

本评测任务所使用的新冠开放知识图谱来源于OpenKG的新冠专题（<http://openkg.cn/group/coronavirus>）。其中包括健康、防控、百科、临床等多个具体数据集。我们将这些数据集整合到一起，同开放领域知识库PKUBASE一起作为问答任务的依据。

#### 输入输出

* **输入**

输入文件包含若干行中文问句。

* **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。

* **输入样例**

q1:抗疫英雄钟南山院士发表过哪些文章？

q2:N95口罩到底哪里好？

q3:新冠病毒的命名组织是什么时候成立的？

* **输出样例**

<慢性阻塞性肺疾病在中国>\t<感冒后咳嗽敏感性及气道神经源性炎症改变>\t<连花清瘟胶囊体外抗甲型流感病毒的实验研究>

"气密性好"\t"过滤效率高"\t"防飞沫"

"1948年4月7日"

* **说明**

为帮助参赛选手提高系统性能，训练文件会提供问题对应的SPARQL查询。

例句一对应的SPARQL查询语句如下：

 SELECT ?x WHERE

 {

<钟南山> <重要成果> ?x.

?x <类型> <文章>.

}

例句二对应的SPARQL查询语句如下：

 SELECT ?x WHERE

{

?y <类型> < N95口罩>.

?y <优点> ?x.

}

例句三对应的SPARQL查询语句如下：

 SELECT ?x WHERE

{

< 2019新型冠状病毒> <命名组织> ?y.

?y <成立时间> ?x.

}

SPARQL语言的语法规则可以参考<https://www.w3.org/TR/rdf-sparql-query/>。原则上不要求生成SPARQL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。**

输出答案中由<>括起的为实体，由""括起的为文本值。请严格按照问题顺序进行答案输出，若参赛系统对某一问题产生答案为空，请输出空行。

#### 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：

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

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

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

#### 数据集

* **数据来源**

本次测评的问答数据来自于人工构建与标注。标注过程不依赖于特定模板，标注人员超过20人以保证问题各方面的多样性。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（对应单元组查询），也包含复杂问题（对应多元组查询），它们的数量比例大致为1:1。

* **训练集&验证集**

在训练数据发布阶段，我们会发布3000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问题作为验证集，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，给出得分并进行排行。

在测试数据发布阶段，我们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问题，作为测试。

* **知识库相关文件说明**

本任务使用的知识库主要由开放领域知识库PKUBASE及OpenKG新冠专题内的各数据集（<http://openkg.cn/group/coronavirus>）融合得到。整合后的用于该任务的知识库相关文件下载地址为：https://pan.baidu.com/s/1MOv9PCTcALVIiodUP4bQ2Q 密码：hcu8。参赛选手可以下载数据文件后使用相应的知识库管理系统（例如gStore系统：<http://gstore-pku.com/> ）进行存储和查询。同时，为方便参赛选手完成任务，我们也提供在线查询终端，选手可以通过浏览器或调用API进行SPARQL查询。详情访问<http://pkubase.gstore-pku.com/>。

数据说明（如有变动后续会进行更新，以实际下载文件为准）：

1. triples.txt包含知识库主要三元组。
2. types.txt包含各实体的类别三元组。
3. schema.txt包含类型和谓词之间的上下位关系等信息。
4. mention2ent.txt可以用来辅助选手进行实体链接。其格式为“短语\t候选实体\t候选实体的排名”，如下例短语“逆时针”排名第一的候选实体是逆时针\_（汉语名词）。此文件仅供参考，其覆盖范围及准确度并无保证，选手可以视情况使用。

逆时针 逆时针\_（汉语名词） 1

逆时针 逆时针\_（张靓颖演唱歌曲） 2

逆时针 逆时针\_（化妆品品牌） 3

…

## 任务提交

本次评测额将采取刷榜方式，各任务验证集发布后，允许参赛队伍多次向平台提交结果，文件命名参考具体任务说明，并以队伍名作为前缀。格式与任务描述中的示例输出相同（注意格式，如对于问答任务需要保留<>或""），排名实时更新。参赛队伍可在评测集发布之前随时上传验证集的计算结果，管理系统会及时更新各队伍的最新排名情况；

测试集发布后，允许参赛队伍多次提交测试集结果文件（每天提交不超过2次）。

**最终提交文件要求：**每一个参赛队需提交的材料如下。

1. 问答任务测试集结果文件，用result.txt命名（UTF-8格式）
2. 相关代码及说明
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）

以上三个文件需在任务提交截止日期前发送至邮箱husen@pku.edu.cn。邮件的标题为：“CCKS-OpenKG-参赛队名称”，例如“CCKS-OpenKG-火箭队”。

代码及其文档需打包成一个文件（tar，zip，gzip，rar等均可），用code.xxx命名，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**。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。

本次评测将依托**biendata平台**（https://biendata.com/）展开，请有意向的参赛队伍关注平台上的竞赛列表。

## 时间安排

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：4月3日

公开排名榜(A)：4月3日 - 7月10日

测试数据发布：7月10日

最终排名榜(B)：7月10日 - 7月20日

最终测试结果：7月20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

## 评测规则

注意，以下通用规则适用于本评测当中的所有子任务。在此基础上，参赛选手还需遵循具体任务的特有规则（参见章节2）。如有冲突，以后者为准。

1. 参赛选手需要提交“参赛队名，队长信息（姓名，邮箱，联系电话），参赛单位名称”等信息，报名方式稍后在评测网站发布。
2. 参赛选手可以选择只参加某个（某些）子任务，也可以选择全部参加，以在比赛平台上的结果提交为准。每个子任务独立计算排名和奖励。
3. 报名截止到测试数据集发布，在测试数据集发布之后，未报名的选手/队伍不能再报名或提交。
4. 每支队伍需指定一名队长，队伍名称不超过15个字符，队伍成员不超过4人。
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，将取消所有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。允许使用外部数据，但该数据必须是公开的，并在提交最终结果时一并提交（如数据过大，需提供下载地址）。
7. 参赛选手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，将取消参赛资格**。
8. 欢迎国内外在校生及社会在职人士参加。比赛组织方成员不可参赛。

## 评测奖励

本测评各子任务独立计算，每个子任务第一名5k，第二名3k，第三名2k，技术创新奖（每个子任务至多一项）5k，同时各子任务排名靠前队伍将获授精美参赛奖牌、证书。

## 组织者

* **总体协调：OpenKG**
* **子任务一 新冠百科知识图谱类型推断**

李志强 东南大学 zqli@seu.edu.cn

孙浩然 东南大学 sunwift@seu.edu.cn

吴天星 东南大学 tianxingwu@seu.edu.cn

漆桂林 东南大学 gqi@seu.edu.cn

* **子任务二 新冠概念图谱的上下位关系预测**

张裕舟 哈尔滨工业大学 yuzhouzhang@ir.hit.edu.cn

佘琪星 哈尔滨工业大学 qxshe@ir.hit.edu.cn

王必聪 哈尔滨工业大学 bcwang@ir.hit.edu.cn

刘铭 哈尔滨工业大学 mliu@ir.hit.edu.cn

秦兵 哈尔滨工业大学 bqin@ir.hit.edu.cn

* **子任务三 新冠科研抗病毒药物图谱的链接预测**

卢栋才 华为云 ludongcai@huawei.com

王鹏 华为云 wangpeng231@huawei.com

陈卓 浙江大学 chenzhuo98@zju.edu.cn

* **子任务四 新冠知识图谱问答**

胡森 北京大学王选计算机研究所 husen@pku.edu.cn

林荫年 北京大学王选计算机研究所 linyinnian@pku.edu.cn

邹磊 北京大学王选计算机研究所 zoulei@pku.edu.cn

许斌 清华大学计算系 xubin@tsinghua.edu.cn

常德杰 妙健康 changdejie@miao.cn

刘邦长 妙健康 liubangchang@miao.cn