**CCKS 2021 技术评测任务书**

# 运营商知识图谱的推理问答

# 任务描述

 **基于知识图谱的问答系统，通过对用户输入query进行语义理解，生成结构化查询语句，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。当前知识图谱问答系统在简单句（单实体单属性）上已经取得比较好的效果，而在约束句：条件约束句、时间约束句，以及推理型问句：比较句、最值句、是否型问句以及问句中带有交集、并集和取反的问句等，其逻辑推理能力还有待提升。**

 **以电信运营商场景为例，比如：“不含彩铃的套餐有哪些？”、“支持长途漫游，价格低于100元的套餐有哪些？”、“神州行B套餐是5G套餐吗”等，这类需要推理的Query目前的问答系统难以回答。此次评测任务，是希望参赛选手基于我们提供的运营商知识图谱，提升在线问答的推理能力。**

* **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: 流量日包的开通方式？**

**q2: 不含彩铃的套餐有哪些？**

**q3: 神州行B套餐是5G套餐吗？**

* **输出样例**

**a1:“KTLLRB”**

**a2:“流量月包\t流量年包”**

**a3:“是”**

# 数据描述

* **数据来源：**

 本次测评的问答数据来自于电信运营商业务真实数据，而不是通过模板生成，并且经过多个业务专家进行人工标注，能够保证数据的准确性和多样性。问答数据集中的问题，包含简单句、约束句和推理句，数量比例大致为4：4：2。

* **训练集&验证集：**

 在训练数据发布阶段，我们会发布5000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。

 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。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。

* **知识库相关文件说明**

 本任务使用的知识库是来源于阿里巴巴和中移在线服务有限公司联合构建，详细地址会在阿里天池平台发布。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. schema.txt包含类型和谓词之间的上下位关系等信息。
2. types.txt包含各实体的类别三元组。
3. triples.txt包含知识库主要三元组。
4. synonyms.txt可以用来辅助选手进行实体识别。

# 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：

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

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

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

# 任务提交

**提交方式：**

本次任务将采取刷榜的方式，分为两个阶段进行：

第一阶段：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（每日最多可上传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。

**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 2021官网）

以上三个文件需在任务提交截止日期前发送至邮箱chengguang.tcg@alibaba-inc.com。邮件的标题为：“CCKS-KBQA-参赛队名称”，例如“CCKS-KBQA-飞天队”。

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。

评测平台：

 本次评测将依托阿里天池平台（https://tianchi.aliyun.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

* 任务征集截止：3月15日
* 任务准备时间：3月20日—4月1日
* 评测任务发布：4月1日
* 报名时间：4月1日—7月15日
* 训练及验证数据发布：4月20日
* 测试数据发布：7月15日
* 提交测试结果：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18日—21日

# 规则

1. 报名方式：用淘宝或阿里云账号登入天池官网，完成个人信息注册，即可报名参赛。本次比赛的参赛对象仅限全日制在校大学生（本科、硕士、博士均可）和企业员工；

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5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，参赛选手允许使用预训练数据（如词向量、字向量等），但是不能用额外的标注数据；

# 任务组织者及联系人

组织者：阿里巴巴达摩院机器智能技术、中移在线服务有限公司

联系人：唐呈光 chengguang.tcg@alibaba-inc.com

 石薇薇 shiweiwei.sww@alibaba-inc.com

 杨春杰 yangchunjie@chinamobile.com

届时会开放钉钉群作为任务讨论平台。

# 奖项安排

一等奖（1名）：15000

二等奖（2名）：10000

三等奖（3名）：5000

创新技术奖（2名）：3000