**开放领域的中文问答任务**

1. **任务描述**

本评测任务为基于开放领域知识库的中文问答，简称COQA （Chinese Open-domain Question Answering）。即对于给定的一句中文问题，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、关系抽取等子任务。这些子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。

**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。

**示例**

输入：

徐峥和黄渤共同出演的电影有哪些？

中国第二长河有多长？

输出：

<人再囧途之泰囧>\t<疯狂的石头>\t<印囧>

"约5464公里"

说明：

例句一对应的SPARQL查询语句如下：

 SELECT ?x WHERE

 {

?x <主演> <徐峥>.

?x <主演> <黄渤>.

?x <类型> <电影>.

}

 例句二对应的SPARQL查询语句如下：

 SELECT ?x WHERE

{

<黄河\_（中国第二长河）> <河长> ?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。

1. **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：

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

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

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

1. **数据集：**

**3.1 数据来源**

本次测评数据主要来自于人工构建与标注。

**3.2 训练集&验证集**

在训练数据发布阶段，我们会发布1200条左右标注好的数据（问题/SPARQL/答案）作为训练数据。同时发布400条左右标注好的数据作为验证集。

在测试数据发布阶段，我们会发布400条左右不含标注结果的问题，作为测试。

**3.3 知识库相关文件说明**

本任务使用PKU BASE作为指定知识图谱。PKU BASE用于该任务的版本下载地址为：https://pan.baidu.com/s/1MOv9PCTcALVIiodUP4bQ2Q 密码：hcu8。参赛选手可以下载数据集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PKU BASE的在线查询终端，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：

1. pkubase-triples.txt包含知识库主要三元组。
2. pkubase-types.txt包含各实体的类别三元组。
3. pkubase-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，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

…

1. **任务提交**

本次任务将采取刷榜的方式，验证集发布后，允许参赛队伍多次向平台提交结果，文件命名为“参赛队名称\_valid\_result.txt”，格式与任务描述中的示例输出相同（注意保留<>或""），排名随时更新。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。

**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 2018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-COQA-参赛队名称”，例如“CCKS-COQA-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。

1. **时间安排**

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

1. **组织者**

北京大学计算机技术研究所，邹磊

北京大学计算机技术研究所，胡森

关于本评测有任何问题，请联系：husen@pku.edu.cn