**CCKS 2021 技术评测任务书**

# 面向中文医疗科普知识的内容理解

# 一、任务描述

随着科技的发展和经济的进步，人们对健康问题越来越关注，医疗科普资讯的普及也得到了国家和社会的一致认可和重视。互联网技术日新月异，图文、问答、短视频和直播等新内容形式层出不穷，如何利用海量的医疗资讯内容快速满足用户需求以及保证内容的准确、权威，都是很重要的课题。

所以，本次评测任务围绕着中文医疗科普知识的内容理解展开，分为两个子任务：

1. **医疗科普知识阅读理解**
2. **医疗科普知识答非所问识别**

**子任务一：医疗科普知识阅读理解**

1. 任务定义

医疗科普知识阅读理解任务针对用户所提出的搜索**query**，在相关的文章中找到对应的**答案**片段内容，以作为直接展示给用户的摘要。

本次任务参考了CMRC2018和CJRC2019阅读理解任务：

1. 拒答类答案定义：文章里面没有能回答query的片段或摘要，人工标注结果为空（0条标注答案）。
2. 单个答案定义：每一条标注答案由人工**独立**标注；单个答案可以由一段连续的文本组成，也可能由文章中若干不连续的短语/句子/片段组成。具体json格式见2.2示例。
3. 参赛者只需对每条测试数据预测出最多1条答案；如果是拒答类，给出空结果即可。
4. 数据描述

**2.1 格式如下：**

**Json**文件格式。

**context\_id**: 文档ID

**context\_text**: 正文内容

**qas**: 问答对列表。其中:

**query\_text**: 问题描述

**query\_id**: 问题ID

**answers**: 答案列表，其中：

每个元素为一个字符串列表，代表一个人工标注答案。

如下图所示，第一个答案来自文章中若干不连续的短语/句子/片段，组成一个字符串列表；第二、三个答案分别是一段连续的文本；每个答案之间相互独立标注。

**2.2 示例**：



**2.3 数据集说明**

本评测子任务的数据集包括：

1）训练及第一阶段测试集：27000条标注数据作为训练集，3000条非标注数据为第一阶段测试集。

2）最终测试集：在测试数据发布阶段，会有6000条非标注数据作为最终测试集。最终测试集分为公域和私域两部分数据（各3000条），公域数据作为闭榜之前的排名依据，参赛队每次需提交全部最终测试集的结果；闭榜后的最终排名使用全部最终测试集的预测结果（6000条）计算得出。

3）非标注数据无**answers**字段

4) 训练集包含0或1个人工标注答案，其中0个标注答案表示‘拒答’类；测试集将包含0~3个标注答案，有多于1个的标注答案表示有多个人工标注，且都是正确答案；计算评测指标时会将预测结果与每个标注答案依次计算，取最高的F1分数作为本条结果的得分。（注：测试集包含多个答案的原因是因为现实场景下每个人对答案的理解会有不同，并且是多人标注，取最高匹配得分）

5）除训练集和测试集外，主办方会提供100条样例标注数据作为验证集（含指标计算脚本）；验证集会包含0~3个标注答案，验证集的评测结果不作为任何排名依据，并且不能加入训练集使用。

1. 评价指标

本次子任务采用F1值来评估预测答案效果，预测数据提交格式同训练集，每条结果包含**0~1**个预测答案。

指标计算脚本metrics.py会随训练集一起发布，公式定义如下：

$$L\_{g}=Len\left(gold\right)$$

$$L\_{p}=Len\left(pred\right)$$

$$L\_{c}=InterSec\left(gold, pred\right)$$

$$precision=\frac{L\_{c}}{L\_{p}}$$

$$recall=\frac{L\_{c}}{L\_{g}}$$

$$f1(gold,pred)=\frac{2\*precision\*recall}{precision+recall}$$

$$F1=\max\_{i}f1(gold\_{i},pred)$$

$$FinalF1=\frac{\sum\_{i=1}^{N}F1\_{i}}{N}$$

**子任务二：医疗科普知识答非所问识别**

1. 任务定义

答非所问是指针对问答形式的医疗科普内容**<Q，A>**，回答**A**和问题**Q**没有关联，即答案不能满足所提问题。在医疗科普内容里，一般有两种答非所问形式，一种是回答**A**和问题**Q**完全没有关联，不能回答提问的问题；第二种是回答**A**和问题**Q**有关联，但回答的方向不对，没有回答提问的问题。

举例如下：

**第一种**：

**Question:** 脸总是起痘是闭口粉刺，还总是泛红怎么治疗呢？

**Description:** 脸总是起痘泛红是闭口粉刺，该用什么药

**Answer:** 皮肤光亮泛红，很可能是由化妆品过敏引起的过敏，如果脸部表皮薄，可能有皮肤冲洗过敏，所以应停止使用化品，然后应用炉甘石洗剂，也可以采取抗过敏药物治疗内部，并避免辛辣刺激性食物，别熬夜，多喝水，没特效药，注意

生，可用扑尔敏试试。

**标注评语：**患者询问治理粉刺方法和用药，回答是关于化妆品过敏的内容，没有明显关联。

**第二种**：

**Question：**气管狭窄是怎么回事？

**Description：**我老公最近胸闷特别厉害，而且咳嗽不断去医院检查，是右上叶支气管狭窄，担心是癌症。右上叶支气管狭窄是癌症吗？

**Answer：**一般良性的狭窄，多数可以考虑用支气管镜下球囊扩张的治疗方法。如果是短暂的狭窄 ，比如气管异物引起的肉芽导致的狭窄，把气管异物通过支气管镜取出来之后，就可以看肉芽很快消失。如果是恶性狭窄，恶性狭窄主要是指肿瘤，这种情况的治疗就更加复杂，除了肿瘤是否能够切除，是否能够外科治疗之外，还要考虑支气管镜的介入治疗。

**标注评语：**患者询问气管狭窄原因，回答是气管狭窄的治疗方法，回答的方向不正确。

1. 数据描述

**2.1 格式如下：**

**Label** \t **Docid** \t **Question** \t **Description** \t **Answer**

**Label**：标注，0/1，其中1表示**Answer**不能回答**Question**，是答非所问。

**Docid**：文档id

**Question**：问题

**Description**：针对问题的补充描述

**Answer**：回答

**2.2 示例**：

**Label**: 1

**Docid**：abcd023659647xigos

**Question**：出肉芽的甲沟炎怎么治疗

**Description**：甲沟炎长了肉芽，大概黄豆粒般大小，目前在用碘伏和红霉素软膏，没什么效果，除了拔甲手术还有什么更好的治疗方法吗？

**Answer**：从你说的情况来看这是局部的支架，长得不好了，再就是有甲沟炎这种情况得先治甲沟炎好了以后再拔甲，如果是这个指甲卡到了肉里边，一定要先把指甲拔下来，不把拔甲，只能是先消炎，以后指甲长出来还会引起甲沟炎的。

**2.3 数据集说明**

本评测子任务的数据集包括：

1. 训练及第一阶段测试集：40000条标注数据作为训练集（正负比例大致1：2），5000条非标注数据为第一阶段测试集，用于在最终测试集发布前榜单排名的依据，提交格式和训练集一致。如需离线优化，可使用训练集做n-fold划分。
2. 最终测试集：在测试数据发布阶段，会有10000条非标注数据作为最终测试集。最终测试集分为公域和私域两部分数据（各5000条），公域数据作为闭榜之前的排名，参赛队每次需提交全部最终测试集的结果；闭榜后的最终排名使用全部最终测试集的预测结果（10000条）计算得出。
3. 标注数据格式为每行一条数据，有5列，用\t分隔，依次为**Label**、**Docid** 、**Question**、**Description**和**Answer**；非标注数据格式为每行一条数据，有4列，用\t分隔，依次为**Docid**、**Question**、**Description**和**Answer**。
4. 评价指标

本次子任务采用精确率（**Precision**, **P**）、召回率（**Recall**, **R**）、**F1**值（F1-measure, F1）来评估答非所问的识别效果。

$$P=\frac{模型预测为答非所问 且 标注为答非所问}{模型预测为答非所问}$$

$$R=\frac{模型预测为答非所问 且 标注为答非所问}{标注为答非所问}$$

$$F1=\frac{2\*P\*R}{P+R}$$

# 二、任务提交

本次任务将采取刷榜的方式，第一阶段测试集发布后，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标注数据格式相同，排名实时更新。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（每日最多可上传3次），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结果文件（每日不超过3次），文件命名为“参赛队名称\_test\_result.txt”，结果文件提交格式同上。

**最终提交文件要求**：

每一支参数队伍提交的材料如下（阅读理解和答非所问识别分开提交）：

1. 结果文件，result.txt（utf-8格式）
2. 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3个文件需在任务提交截止日期前发送至指定邮箱。

阅读理解：发送至ccks2021Medical\_1@qq.com，邮件标题格式为：CCKS-评测任务-MRC-最终提交文件-参赛队名称”，例如“CCKS-评测任务-MRC-最终提交文件-医心队”。

答非所问：发送至ccks2021Medical\_2@qq.com，邮件标题格式为：CCKS-评测任务-WA-最终提交文件-参赛队名称”，例如“CCKS-评测任务-WA-最终提交文件-医心队”。

代码及其文档需打包成一个zip文件，用code.xxx命名，要求提交所有的程序代码及相关的配置说明，确保程序能够正确运行，且所得结果与result.txt相符。

方法描述文档用desc.pdf命名，包含算法描述及参数设置，需用pdf格式存储，页数不超过10页。

# 三、规则

本次比赛组委会依托Biendata（<https://www.biendata.com/>）作为在线评测平台，所有参赛队伍最终结果均需在Biendata上指定的时间窗口内进行提交。

1. 参赛选手须在Biendata平台管理系统中注册，并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；
2. 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（包括队长），报名截止日期之后不允许更改队员名单；
3. 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；
4. 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；
5. 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；除主办方提供的数据集外，参赛选手允许使用公开的预训练模型（如词向量、字向量等），但是不能用额外的标注数据。
6. 参赛队伍可在参赛期间随时上传验证集的预测结果，一天不能超过3次 ，管理系统会实时更新各队伍的最新排名情况。
7. 参赛队伍最终需要提交可运行的代码和方法描述文档，如果在排行榜上的结果无法复现，将取消参赛资格。
8. 欢迎在校师生及企业在职人士参赛。

# 三、时间安排

时间安排初定如下，如有后续调整，将在讨论群中更新说明。

* 评测任务发布：4月1日
* 报名时间：4月1日—7月15日
* 训练及第一阶段测试集数据发布：4月20日
* 第一阶段测试集提交开放：4月20日 - 7月15日
* 最终测试集数据发布：7月15日
* 最终测试集提交开放： 7月15日- 7月20日
* 评测论文提交：8月5日
* CCKS会议日期：8月18日—21日

# 四、奖励设置

第一名：人民币15000元

第二名：人民币7000元

第三名：人民币3000元

技术创新奖：人民币5000元

备注：每个子任务独立设置奖项，包括一、二、三名和技术创新奖。子任务每个奖项最多一支队伍获得，技术创新奖可以缺失，并且可以和前三名兼得；参赛队伍可以分别参加两个子任务，并分别排名；获奖者有机会进入2022年腾讯医疗校园招聘终面和实习生绿色通道。

# 五、任务组织者及联系人

刘新春（腾讯）：xinchunliu@tencent.com

吴贤（腾讯）：kevinxwu@tencent.com

陈裕通（腾讯）：bradleychen@tencent.com

邵纪春（腾讯）：sunnyshao@tencent.com

周佳（腾讯）：alisazhou@tencent.com

# 六、任务交流群组

任务交流平台采用微信群的方式。所有报名参赛的参赛队要求至少有一名成员加入到该讨论群中，后续所有有关本任务评测及数据的相关说明和通知将只在讨论群中发布和交流，不再另行发邮件或更新官网通知，请所有参赛队务必在报名成功之后按子任务发申请邮件到以下邮箱申请进群，邮件头格式：申请加入讨论群：真实姓名-参数队名-Biendata注册名-微信号。

阅读理解子任务：ccks2021Medical\_1@qq.com

答非所问子任务：ccks2021Medical\_2@qq.com