

# MDS-1 – 数据标注规范

---

## 一、数据格式：

u0, r1, u1, r2.....rn

一段对话由多个句子组成，其中有两种角色，分别是病人和医生。其中病人和医生的对话并不是简单的交替进行的，其中一人可以连续多次述说语句。

## 二、任务介绍：

将对话中的每句话做好标记，标记主要由两部分组成，一部分是意图Intent，另一部分是Aspect-Specific Aspect-Value对。Aspect是一个大的方面，例如疾病，症状，药物等，每个Aspect会有多个具体方面，每个具体方面可能会有值，例如症状的包含咳嗽，发烧等具体方面。而发烧这个的具体方面的Value可以是具体的温度。（每一句话至少要标注Intent,其余的Aspect-Specific Aspect-Value均为非必需，Value可能很多对话均不需要标注，可为空）

标注时尽量选择完整的一句或者子句进行标注，要满足语义的完整性。例如：“你好，这种情况多久了呢？”应该全部选择然后进行标注，而不是仅选择“多久”。但是明显多余的无用信息也不用全部标注，例如“干扰素和解霉素治疗病毒感染吗？（，）咨询人性别:女， 咨询人年龄：1-7岁”这句话就可以仅对“干扰素和解霉素治疗病毒感染吗”进行标注，后面的就不标注了。每段对话第一句的病人的年龄，性别信息不用标注。

### Intent

- 1) **inform (告知)**，用于告知有关信息（一般包含Aspect和Specific Aspect，可能不含Value）
- 2) **inquire (询问)**，用于向对方提问，例如“是不是感冒了？”给出Aspect-Specific Aspect对猜测的问题或者或者“生的什么病？”等并没有给出Specific Aspect只给出了Aspect的开放式问题。
- 3) **qa (问答)**，用于医疗知识的问题或者回答关于医疗知识的提问。这类句子与当前病情可能相关或不相关，但是与诊断过程无关，常见的诊断过程一般为病人告知症状，发病时间，部位等信息，医生做出诊断，给出疾病，或者服用药物等信息。（只标记Intent）。
- 4) **diagnosis (诊断疾病)**，用于做出诊断（后面的Aspect只能选择疾病）。
- 5) **recommend (推荐)**，用于向病人给出推荐（后面的Aspect可以选择医疗机构/药品/治疗方法等）
- 6) **chitchat (闲聊)**，用于礼貌用语等（只标注Intent）。
- 7) **other (其他)**，用于其他未指明的情况（只标注Intent）。

## 三、例子说明：(用P表示病人，D表示医生)

P:发烧（女，1岁7个月）	Intent=Inform Aspect=symptom Specific Aspect=发烧
D:你好，请问现在孩子体温多少	Intent=Inquire Aspect=temperature Specific Aspect=体温
P: <u>39度5</u>	Intent=Inform Aspect=temperature Specific Aspect=体温 Value=39.5

## 四、注意事项：

1.一句话中可能会有多个Intent或者Intent，都要标记，有的句子例如“好的”加其他的信息，这里的好吧就可以不用在意。而如果有一句的单独的好吧，就可以标为Chitchat。**就是要找准句子的主要内容。尽量标记完整的一句或者子句，明显的不同Intent可以进行拆分。**

case1：

P: <u>最近头痛</u> ， <u>得什么病了</u> 啊， <u>应该吃点什么药</u> 啊？	Intent=Inform Aspect=symptom Specific Aspect=头痛 Intent=Inquire Aspect=disease Intent=Inquire Aspect=medicine
D: <u>你还有什么其他的症状</u> 吗？	Intent=Inquire Aspect=symptom

case2:

D: <u>你可以吃布洛芬</u>	Intent=Inform Aspect=medicine Specific Aspect=布洛芬
P:好的， <u>请问一天吃几次</u>	Intent=Inquire Aspect=frequency
D: <u>一天三次</u>	Intent=Inform Aspect=frequency Specific Aspect=null Value=一天三次
P: <u>好的</u>	Intent=Chitchat (Aspect 和Specific Aspect都不需要)

2.有些提问属于问答的范畴，只需要标记下病人Intent=Qa就可以，医生的回答也直接用Qa标记。

case：

D:请问什么是血糖	Intent=Qa
P:血糖就是血液中葡萄糖的含量	Intent=Qa

3.对于某些Aspect，可能很多未必有明确的回复，可能会出现不知道或者不清楚这样的回答。这里可以统一将其定义为not-sure。

case：

D:请问你青霉素过敏吗？	Intent=Inquire Aspect=symptom Specific Aspect=青霉素过敏
P:不清楚	Intent=Confirm Aspect=symptom Aspect=青霉素过敏 Value=not_sure

4.表达是什么或者是不是的提问都可以看做是Inquire，例如有“什么症状”，“是不是感冒了？”

case:

P:是不是感冒了?	Intent=Inquire Aspect=symptom Specific Aspect=感冒
-----------	---

D:有什么症状?	Intent=Inquire Aspect=symptom
----------	-------------------------------

P:感冒吃什么药?	Intent=Inquire Aspect=medicine Specific Aspect=感冒
-----------	--

5.有时候语句中出现的内容比较偏口语化的表达，或者与Aspect的Value值不会完全相同，需要大家自己判断寻找下最匹配的Value进行标记。

case:

P:现在就是时不时的发烧	Intent=Inform Aspect=symptom Specific Aspect=反复发烧
--------------	--

6.对于一些出现在句子里可以标记Aspect，但是Value在所给的信息中没有的可以自己添加Value，但是添加的Value一定要是从原句子中复制的片段。

case:

D:发生的频繁吗?	Intent=Inquire Aspect=frequency
P:上午少，晚上多。	Intent=Inform Aspect=frequency Specific Aspect=null Value=上午少，晚上多（Value中本没有这个选项，标记时自行添加）

7.存在某些限定情况下的不好标注的就省略掉部分信息。

case:

D：一个礼拜后好不了再去医院做CT	Intent=Inform Aspect=check_item Specific Aspect=CT(省略掉前面的时间限定条件)
-------------------	---

8. 存在某些需要从前文（不仅仅是相邻的前文，相隔几句也是可以的）获取信息补全Specific Aspect的例子。

case:

P:阿莫西林胶囊怎么吃?	
D:一次两粒	Intent=Inform Aspect=frequency Specific Aspect=阿莫西林 Value=一次三片

## 五、操作流程：

1. 打开index.html网页，在网页顶部点击打开对话并选择对话文件，可以看到页面左侧的对话。



2. 单击选中一条对话，可以在页面右侧看到标注栏。（用鼠标选择句子中完整的一句，排除较为明显的无效信息。在原始消息框中选择语句中对应的部分点击“添加意图(add Intent)”。（例子如下，只选择了有效的前半句，后半句不用选择）

原始消息 (Initial Message) 干扰素和解毒素治疗病毒感染吗 (,) 咨询人性别: 女咨询人年龄: 1-7岁

+ 添加意图(Add Intent)

x

语句(Sentence)

干扰素和解毒素治疗病毒感染吗

下标(Index) 0 13

意图(Intent) 例如: 告知, 询问, 诊断。 e.g.: Inform, Inquire, diagnosis.

QueryAnswer(问答)

方面(Aspect) 例如: 症状, 疾病, 药品。 e.g.: Symptom, Disease, Medicine.

选择或填写 slot...

具体方面(Specific Aspect) 例如: 方面: 药品; 具体方面: 布洛芬, 地塞米松等。 e.g.: Aspect: Medicine; Specific Aspect: Ibuprofen, Dexamethasone.

选择或填写 aspect...

具体方面的值(Value) 例如: 方面: 频率; 具体方面: 布洛芬; 具体的值: 一天三次。 e.g.: Aspect: Frequency; Specific Aspect: Ibuprofen; Value: Three Times A Day.

Select

这里可以认为是与病情上下文无关的咨询，于是标记为QuestionAnswering。

如果标注出现错误，可以单击左上角的“x”删去这一标注，重新点击添加重新进行标注  
标注完一句之后在左侧点击下一句，或是单击标注页面的下一条进行标注

×

语句(Sentence)

确诊脚部病毒感染啦?

下标(Index)09

意图(Intent) 例如: 告知, 询问, 诊断。 e.g.:Inform,Inquire,diagnosis.

Inquire(询问, 用于询问, 例如“是不是感冒了”, “有什么症状啊”等各种疑问问题)

方面(Aspect) 例如: 症状, 疾病, 药品。 e.g.:Symptom,Disease,Medicine.

Disease(疾病)

具体方面(Specific Aspect) 例如: 方面: 药品; 具体方面: 布洛芬, 地塞米松等。 \ne.g.:Aspect:Medicine;Specific Aspect:Ibuprofen,Dexamethasone.

脚部病毒感染

具体方面的值(Value) 例如: 方面: 频率; 具体方面: 布洛芬; 具体的值: 一天三次。 \ne.g.:Aspect:Frequency;Specific Aspect:Ibuprofen;Value:Three Times A Day.

是

当句子中出现同种意图方面, 多个具体方面时, 不同的具体反复用中文逗号隔开, 例如推荐了多种注意事项: (分成多次进行标记, 询问了两次药物也是两次的Intent)

生活方式调整, 生活规律, 避免烟酒辛辣食物刺激不要熬夜久坐, 多运动锻炼身体, 多食坚果核桃南瓜籽花生之类的食物。心理疏导也很关键

+ 添加Action

×

语句(Sentence)

生活方式调整, 生活规律, 避免烟酒辛辣食物刺激不要熬夜久坐, 多运动锻炼身体, 多食坚果核桃南瓜籽花生之类的食物。心理疏导也很关键

下标(Index)061

意图(Intent) 例如: 告知, 询问, 诊断。 e.g.:Inform,Inquire,diagnosis.

Recommend(推荐药品/治疗方法/检查项目/医疗位置/注意事项/药物种类/科室, 例如“推荐吃一些布洛芬。”)

方面(Aspect) 例如: 症状, 疾病, 药品。 e.g.:Symptom,Disease,Medicine.

Precaution(注意事项)

具体方面(Specific Aspect) 例如: 方面: 药品; 具体方面: 布洛芬, 地塞米松等。 e.g.:Aspect:Medicine;Specific Aspect:Ibuprofen,Dexamethasone.

拒绝辛辣食物, 勿熬夜久坐, 多食坚果核桃南瓜籽花生类食物, 心理疏导

具体方面的值(Value) 例如: 方面: 频率; 具体方面: 布洛芬; 具体的值: 一天三次。 e.g.:Aspect:Frequency;Specific Aspect:Ibuprofen;Value:Three Times A Day.

填写value...

3. 选择Aspect后, 部分Aspect的Specific Aspect需要输入关键字进行检索, 另一部分会出现候选项。如果对话中的Specific Aspect不在系统中存在的情况下可以自己填写Specific Aspect (优先选择系统中的进行填写, 自己填写时请尽量用原文的片段, 例子如下图中的脚部病毒感染)

原始消息  
(Initial Message)

确诊脚部病毒感染啦?

+ 添加意图(Add Intent)

x

语句(Sentence)

确诊脚部病毒感染啦?

下标(Index)09

意图(Intent) 例如：告知，询问，诊断。 e.g.:Inform,Inquire,diagnosis.

Inquire(询问，用于询问，例如“是不是感冒了”，“有什么症状啊”等各种疑问问题)

方面(Aspect) 例如：症状，疾病，药品。 e.g.:Symptom,Disease,Medicine.

Disease(疾病)

具体方面(Specific Aspect) 例如：方面：药品；具体方面：布洛芬，地塞米松等。 e.g.:Aspect.Medicine;Specific Aspect.Ibuprofen,Dexamethasone.

脚部病毒感染

具体方面的值(Value) 例如：方面：频率；具体方面：布洛芬；具体的值：一天三次。 e.g.:Aspect.Frequency;Specific Aspect.Ibuprofen;Value:Three Times A Day.

Select

4. 标注完成后可以通过页面上方的“**下载已标注文件**”保存你的工作成果。