

CMelE-V2增补标注规范

标注细则：

1. 采用篇章标注的方式对语料进行标注，为便于训练，构建句子语料时，主要以句号（也包括叹号，问号等，分号，但语料中较少涉及）为隔断，对语料进行分句；

2. 标注语料没有给出实体的具体下标，如果中心词会出现多次，根据语料的实际语境作为中心词；如果主题词存在医学上的逻辑关系也可以进行标注；

例子：登革热@大90%的登革出血热 (dengue haemorrhagic fever, DHF) 病例为5岁以下儿童。登革热@典型的登革热更常见于成人，而非儿童。

标注：

<登革出血热，多发群体，5岁以下的儿童>

<登革热，多发群体，成人>

<登革出血热，同义词，dengue haemorrhagic fever>

<登革出血热，同义词，DHF>

解释：首先应该根据实际语境的意思，则应取主题疾病的相关关系，登革热对应的多发群体是成人，登革出血热对应的多发群体是5岁以下的儿童，以及登革出血热和登革热也可以标注为鉴别诊断的关系或者是病理分型的关系，这属于医学逻辑的关系,可以标注也可以不标注

3. 实体名称是出现了多次的，也就是相同的实体在不同句子多次出现；遵循就近或按照语义就近原则（先找同一个句子里面的，句子里面找不到的寻找距离最近的）；

例子：急性胰腺炎@伴感染性胰腺坏死 — 附加 — 实施规模更大的引流方法、窦道坏死组织清除术或手术坏死组织清除术针对经选择患者组部分患者的治疗推荐对于经皮引流无效的患者，可能需要采用“渐进”方案，实施规模更大或不同方式的经皮引流（例如腔道内镜引流）、“窦道”坏死组织清除术或手术坏死组织清除术。急性胰腺炎@方案实施的时机以及选择需要咨询专科医生。

标注：

<急性胰腺炎，并发症，感染性胰腺坏死>

<急性胰腺炎，手术治疗，引流方法>

<急性胰腺炎，手术治疗，**窦道坏死组织清除术**>

<急性胰腺炎，手术治疗，**手术坏死组织清除术**>

<急性胰腺炎，手术治疗，**腔道内镜引流**>

解释：急性胰腺炎出现了两次，根据语义更近的原则所以选择第一次出现的**急性胰腺炎**，因为和手术治疗在一个句子里面。急性胰腺炎和感染性胰腺坏死两者是医学上的并发症的关系，可以选择标注或者不标注，语义上没有表达出来。

文中表达的意思是急性胰腺炎可以选择引流方法，对于特殊患者选择经皮引流，因为有更具体的手术名称，**腔道内镜引流**，所以我们标注<急性胰腺炎，手术治疗，**腔道内镜引流**>，而“**窦道**”**坏死组织清除术**或**手术坏死组织清除术**这个手术名称是重复出现的，我们选择实体之间最近的即可，即选择和**急性胰腺炎**同一个句子里面的**窦道坏死组织清除术**或**手术坏死组织清除术**进行标注。

4.实体范围按照最大范围原则标注实体之间是不允许嵌套，注意**症状**的划分，一般给检查/检验和检查/检验结果异常的统一归为**症状**；

例子：第六节心房颤动心房颤动（简称房颤）在儿童少见，在婴儿罕见。心房颤动@房颤偶尔也可见于心房内手术后，**继发于左心房室瓣关闭不全的左心房扩大**，**WPW综合征**等。

标注：

<心房颤动，同义词，房颤>

<房颤，手术治疗，心房内手术>

<左心房室瓣关闭不全的左心房扩大，相关（导致），房颤>

<wpw综合征，相关（导致），房颤>

解释：左心房室瓣关闭不全对左心房扩大作修饰信息，所以最大化标注实体。

5.复杂语义关系的比如包含多层的关系的标注案例，如果是多层关系的格式，则需要标注实体和最外层的关系即可，不需要实体和内层的实体再次关联。如果外层关系是泛化词这种情况下可以标注内层的更具体的情况。

例子：一种特殊的恶性室性心律失常-尖端扭转型室速可发生于QT间期延长综合征（long QT interval syndrome, LQTS）患儿，并可引起晕厥与猝死。药物可直接（**terfenadine**，**阿司咪唑**）或通过抑制代谢（**红霉素**，**酮康唑**）来延长QT间期。

标注：

<QT间期延长综合征，药物治疗，**terfenadine**>

<QT间期延长综合征，药物治疗，**阿司咪唑**>

<QT间期延长综合征，药物治疗，红霉素>

<QT间期延长综合征，药物治疗，酮康唑>

<QT间期延长综合征，同义词，long QT interval syndrome>

<QT间期延长综合征，同义词，LQTS>

<QT间期延长综合征，相关（导致），晕厥>

<QT间期延长综合征，相关（导致），猝死>

解释：药物可直接（terfenadine，阿司咪唑）或通过抑制代谢（红霉素，酮康唑）来延长QT间期，这里的"药物"是个泛化词，我们就选择更具体的药物名称进行了标注

6.同一个实体和不同的实体构建关系的时候，语境不同实体类型可能也不同；

例子：心肌炎@多巴酚丁胺和米力农会增加室性心律失常风险

标注：

<室性心律失常风险，相关（导致），室性心律失常>

<室性心律失常风险，高危因素，多巴酚丁胺>

<室性心律失常风险，高危因素，米力农>

解释：多巴酚丁胺和米力农是常规的实体类型应该是药物，但是根据该句的语境药物多巴酚丁胺和米力农是引发室性心律失常的因素，因此这里把这两个药物和室性心律失常的关系标注成是高危因素关系最合适，当标注为高危因素的关系时候，多巴酚丁胺和米力农的实体类别对应的是社会学。

注：高位因素关系类型定义：{"subject_type": "疾病", "predicate": "高危因素", "object_type": "社会学"}

7.实体（实体的英文名字，缩写）如果这样的格式，可以某个实体分别和英文名字和缩写对应给到同义词关系即可。实体英文名字和缩写不需要给同义词关系；

例子：苯丙酮尿症(phenylketonuria,PKU)是一种常染色体隐性遗传疾病，因苯丙氨酸羟化酶基因突变导致酶活性降低，苯丙氨酸及其代谢产物在体内蓄积导致疾病。

标注：

<苯丙酮尿症,同义词，phenylketonuria>

<苯丙酮尿症,同义词，PKU>

解释：疾病和自己的别名和缩写分别建立同义词关系即可，别名和缩写之间不需要再建立同义词关系，同理其他关系只需要和主题疾病建立关系即可，不需要和其别名和缩写分别建立关系，文本里

面的只需要和"苯丙酮尿症"建立其他临床表现关系即可

8.几种较难区分的关系：

{"subject_type": "疾病", "predicate": "并发症", "object_type": "疾病"}

{"subject_type": "疾病", "predicate": "相关（导致）", "object_type": "疾病"}

{"subject_type": "疾病", "predicate": "相关（转化）", "object_type": "疾病"}

{"subject_type": "疾病", "predicate": "相关（症状）", "object_type": "疾病"}

难点区分：

并发症：

一种是指一种疾病在发展过程中引起另一种疾病或症状的发生，后者即为前者的并发症。另一种是指在诊疗护理过程中，病人由患一种疾病合并发生了与这种疾病有关的另一种或几种疾病。一般具有三个特征：

①后一种疾病的发生是由前一种疾病所引起的

②从后一种疾病的发生规律上看，前后疾病之间不具有必然的因果关系，只具有偶然的因果关系。因此，后种疾病的出现属突发性。

③后种疾病的出现非因医务人员的过失所致

标注样例1：脑膜炎是流行性腮腺炎儿童期最常见的并发症，常在腮腺炎高峰时出现。

<流行性腮腺炎儿童期，并发症，脑膜炎>

相关（导致）

指一种疾病的发生导致了另一种疾病，或一种疾病继发于另一种疾病。该疾病与其导致的疾病同时存在。疾病的相关（导致）要注意前后关系，在标注过程中，把病因疾病放在前面，导致的疾病放在后面。其中，病因明确为疾病时，关系为相关（导致）。相关导致和并发症比较难区分，一般以当前的语义为准。

标注样例1：假丝酵母菌病是由假丝酵母菌属引起的皮肤、黏膜、脏器的急性、亚急性或慢性炎症，少数可引起脓毒症。

<假丝酵母菌病，相关（导致），脓毒症>

相关（转化）

指一种疾病转化为另一种疾病。转化即原来的疾病消失转而成为另外一种疾病。

标注样例1：在农村、内地、边远地区仍多见由感染性肺炎发展为重症肺炎呼吸衰竭。

<感染性肺炎，相关（转化），重症肺炎呼吸衰竭>

相关（症状）

指一种疾病的症状表现为另一种疾病的症状；或一种疾病表现的症状为疾病。

标注样例1：新生儿肺炎起病前常有上呼吸道感染的症状。

<新生儿肺炎，相关（症状）,上呼吸道感染>