

标注规范

您将看到一段医生和患者的对话文本，包含疾病、患者提问和对话内容。您需要标注出以下三个内容：每句话所对应的话语行为、句子内部记录（包括实体识别和实体规范化）、每个病例的诊疗报告。具体说明如下：

（一）话语行为标注

每句话都有其对应的话语行为（dialogue act），话语行为一共包含十六类，分别是提问-症状、告知-症状、提问-病因、告知-病因、提问-基本信息、告知-基本信息、提问-已有检查和治疗、告知-已有检查和治疗、提问-用药建议、告知-用药建议、提问-就医建议、告知-就医建议、提问-注意事项、告知-注意事项、诊断、其他。

- （1）提问-症状：与症状有关的提问；
- （2）告知-症状：与症状有关的回答，也包括患者的肯定或否定回答；
- （3）提问-病因：询问症状和疾病产生的原因；
- （4）告知-病因：解释症状和疾病产生的原因，通常有“导致”、“引起”、“造成”等词；
- （5）提问-基本信息：基本信息包括患者个人信息、生活习惯（饮食习惯等）、既往病史、家族情况等；
- （6）告知-基本信息
- （7）提问-已有检查和治疗：已有检查和治疗包括做过的检查项目、化验单情况和药物治疗等，该动作必须由医生发出；
- （8）告知-已有检查和治疗：该动作必须由患者发出；
- （9）提问-用药建议：与药物及用药方式相关的提问；
- （10）告知-用药建议：与药物及用药方式相关的回答；
- （11）提问-就医建议：与检查、就医建议相关的提问；
- （12）告知-就医建议：与检查、就医建议相关的回答；
- （13）提问-注意事项
- （14）告知-注意事项
- （15）诊断：①重要信息，通常包含“考虑是”、“应该考虑”、“认为是”等词语，例如“你这个情况需要考虑是消化不良”；②医生通过已有检查报告等信息给出的判断也归为此类，例如“白细胞偏高”。
- （16）其他：①问候、感谢、与病情无关的句子；②空句子，表明该句为图片；③

部分条件语句表示后续的判断，例如“如果查出来有炎症，就需要进行抗感染治疗”；④医生的一些解释，例如“咳嗽是人体清除呼吸道内的分泌物或异物的保护性呼吸反射动作”。

请注意：

- 如果出现两者以上的话语行为，则根据实际情况判断主要的话语行为，无法判断重要性时，则根据就近原则选择第一个分句的话语行为。
- 如果患者提问“该怎么办”、“怎么治疗”，无法明确具体的提问目的，根据上下文选择。
- 如果不确定动作类别，可以结合上下文进行判断，如果发现仍然不属于上述类别，或者对诊疗结果没有影响，可归为其他。
- 有些能反映症状的基本信息需要归为症状类，例如疾病为“小儿消化不良”时，“每天大便几次？”的话语行为是“提问-症状”。

（二）句子内部记录标注

句子内部记录标注分为两大部分：**实体识别以及实体规范化**。

1、实体识别

实体识别是指识别出句子中重要的词组，如症状、药品名、药物类别、服用方式、检查、操作、注意事项等，并对每句话的每个字打上相应的 BIO 标签（B：实体开始，I：实体中间，O：其他文字），需要识别的实体一共包含七类：

- （1）**症状**：指具体症状的名词，例如“咳嗽”、“流鼻涕”、“拉肚子”等；
- （2）**药品名**：指具体的药品名称，例如“我建议你服用妈咪爱”，其中“妈咪爱”就是药品名；
- （3）**药物类别**：需与药品名区分，指一类药物，例如“抗菌药”、“抑制胃酸分泌药”等；
- （4）**服用方式等**：包括用药剂量、用药频次、口服、饭前饭后等相关文字；
- （5）**检查**：与检查建议相关的名词，既包括具体的检查项目，如血常规、尿常规、化验、胸片、支原体检查等，也包括宽泛的建议，如检查、检测等；
- （6）**操作**：与医疗操作相关的名词，例如输液、雾化、挂水等；
- （7）**注意事项**：如喝热水、少食多餐等。

2、实体规范化

实体规范化是针对“**症状**”以及“**检查**”相关的实体提出的。具体说明如下：

(1) 规范化“症状”相关的实体时，在症状实体开头（2:B-症状）的下一行写上规范化名词（只需写在一个表格中即可），例如原文中“流鼻涕”应该规范化为“鼻流涕”。

(2) 规范化“检查”相关的实体时，在检查实体开头（6:B-检查）的下一行写上规范化名词（只需写在一个表格中即可），例如原文中“小便常规”应该规范化为“尿常规”。

请注意：

- 初始化的 BIO 标签已包含了部分信息，包括**机器自动匹配的检查、药品名和检查名**（这部分不准确，需要进一步核实修改，尤其是将药物类别从药品名中分离开）。
- BIO 标注时要尽量标注出关键词，不标无关的词。
- BIO 标签对应如下：0：O，1：I，2：B-症状，3：B-药品名，4：B-药物类别，5：B-服用方式等，6：B-检查，7：B-操作，8：B-注意事项。

(三) 诊疗报告 report

每个病例需要输出一个诊疗报告，参考框架如下：

- (1) 主诉： 主要症状或体征
- (2) 现病史： 主要症状的描述（发病情况，发病时间）
- (3) 辅助检查： 病人已有的检查项目、检查结果、会诊记录等
- (4) 既往史： 既往的健康状况、过去曾经患过的疾病等
- (5) 诊断： 对疾病的诊断
- (6) 建议： 检查建议、药物治疗、注意事项