

# 第三届知识图谱大赛信息抽取测评说明

---

## 一、评估数据集概述

## 二、信息抽取评分

## 三、综合评分

## 四、测试集提交样例

## 一、评估数据集概述

IE\_data.CSV 是一份用于信息抽取任务的数据集，包含两列：

- id：唯一编号，标识每条文本记录。
- text：文本内容，需从中抽取出指定的实体及其类型。

选手需要基于提供的数据集进行信息抽取，将每个文本中的实体与对应的类型进行标注，并生成提交文件。选手提交的数据应包含以下四列：

- id：对应文本记录的唯一编号。
- text：原始文本内容。
- 实体：从文本中抽取的实体。
- 类型：实体所属的类型。

每个抽取的实体及其类型应占据单独的一行。

## 二、信息抽取评分

为了评估选手的信息抽取结果与标准答案之间的匹配情况，本次测评采用 F1 作为主要评价指标。

**F1 指标**F1 值是精确率（Precision）与召回率（Recall）的调和平均数，能够全面评估抽取实体的准确性和覆盖率。计算公式如下：

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- 精确率（Precision）：抽取出的实体中，正确的实体占比。
- 召回率（Recall）：标准答案中的实体被正确抽取的占比。

### 三、综合评分

最终评分将基于选手在所有测试样本上的 F1 值取平均值，作为整体性能的衡量标准。

### 四、测试集提交样例

提交文件应包含四列，分别是：

- id：文本记录的唯一编号。
- text：原始文本内容。
- 实体：抽取出的实体。
- 类型：实体所属的类型。

样例如下（答案仅做演示使用）：

id	text	实体	类型
2	患者,男,46岁,初诊日期 2006年7月6日。主诉: 左侧胸壁疼痛4个多 月。（样例省略）	带状疱疹后遗神经痛	疾病
2	患者,男,46岁,初诊日期 2006年7月6日。主诉: 左侧胸壁疼痛4个多 月。（样例省略）	左侧胸壁疼痛	临床表现
3	患者，女，82岁，于 2013年7月25日， 因“反复双足起疹、瘙 痒数十年，加重1个 月，（样例省略）	双足起疹	临床表现

- 提交形式：以“IE\_队长名\_单位”命名，以csv形式提交，如：IE\_张三\_中国中医科学院.csv。
- 不可改变id列与题目的对应关系。