**一、任务设定：**

 本次竞赛要求参赛者给出一个算法或模型，从评价性文本片段数据集中，结合上下文提取文本片段中的核心实体，即识别文本片段所讨论最重要的实体。竞赛数据由若干文本片段组成，每个文本片段单独占一行，可能包含多个子句，每个子句以半角句号结尾，要求参赛选手识别出每个文本片段子句及其核心实体。

输入输出说明：

输入：逐行读取文本片段；

输出：逐行输出该文本片段的子句以及提取到的核心实体，以json形式分装，格式如下：

[

 {

        "content": "子句1",

        "core\_entity": [

            "entity1",

            "entity2"

        ]

    },

    {

        "content": "子句2",

        "core\_entity": [

            "entity1",

            "entity2"

        ]

    }

  ]

**样例数据：**

文本片段：  
不得不说锤子手机在很多功能操作上的优化真的很用心，尤其是一些看上去并没有什么卵用但让人感觉确实舒服的小设计。

如果说苹果的设计是强行的改变用户习惯，我宁愿选择这个能够随时迁就用户个性的锤子。

答案：  
 [

    {

        "content": "不得不说锤子手机在很多功能操作上的优化真的很用心，尤其是一些 看上去并没有什么卵用但让人感觉确实舒服的小设计。",

        "core\_entity": [

            "锤子手机"

        ]

    },

    {

        "content": "如果说苹果的设计是强行的改变用户习惯，我宁愿选择这个能够随时 迁就用户个性的锤子。",

        "core\_entity": [

            "苹果",

            "锤子"

        ]

    }

]

**二、评分标准：**

 正确识别文本片段中至少一个核心实体得1分，如文本片段包含多个核心实体，每多正确识别一个加0.5分，识别错误反扣0.5直到该子句得分为0为止，实体名称以该实体首次出现在文本片段中的名称为准。

**三、竞赛阶段**

  **训练阶段（4-8月）:**

参赛者获得训练数据，开始训练模型；

发布1.2万条语料片段的训练数据集；

  **测试阶段（8-9月）：**

参赛者获得测试数据，得到模型运行结果后，在线提交，实时刷新测试成绩；

发布20万条语料测试数据集；

  **验证阶段（9月）：**

排行榜前10名队伍上传源代码及说明文档，竞赛评委组使用官方验证数据集进行测试，最终给出排名；

发布20万条语料验证数据集；

注：

1、参赛者可以任意使用其他公开的外部数据源或者词典作为辅助数据，但提交时需要说明清楚使用方式并将使用到的外部数据一并提交，同 时禁止人工标注数据用于系统优化，如有发现成绩归零。

2、 说明文档至少包含编译环境说明，算法、代码等说明。

   3、报名截止日期：8月10日。