

2017 年第二届全国网络舆情分析技术邀请赛

评测大纲

事件关键元素识别

一、任务说明

本任务的目标是从给定语料中抽取出描述指定事件的关键元素。任务的输入为一系列专题事件以及与这些事件对应的文档集合，要求针对每个事件从其对应文档中抽取事件关键元素（注意并非每篇文档中的全部实体），关键元素类型包括时间、地点、参与人物及参与组织等实体，语料中不同事件的新闻数量比例严重倾斜。

二、评测数据集

主办方提供：新闻调试集、新闻调试标注集和新闻测试集。

1. 新闻调试集

新闻调试集包含 100 篇左右新闻报道，共涉及 10 个新闻事件，用于调试参赛系统。

2. 新闻调试标注集

新闻调试标注集包含新闻调试集中每篇新闻报道的关键元素列表，用于调试参赛系统。

3. 新闻测试集

新闻测试集共包含 2 万多篇新闻报道，分为 3 份。每一份测试数据的评测分值占最终评测分值的 10%、30%、60%。

三、参评系统输入输出文件格式

1. 输入文件格式

```
<?xml version="1.0" encoding="UTF-8"?>
<Samples>
  <Sample>
    <SampleID>1</SampleID><!--样本编号-->
    <EventID>1</EventID><!--事件编号-->
    <SampleTitle>河北官方：尽快做好特赦工作 不错放不漏赦</SampleTitle><!--样本标题-->
    <SampleContent><!--样本内容-->
      8月30日上午，河北省召开全省特赦部分服刑罪犯电视电话会议，省委政法委常务副书记王立山要求全省各级各
      有关部门充分认识这次特赦的重大意义，认真学习全国人大常委会《关于特赦部分服刑罪犯的决定》和中央政
      法部门制定的实施办法，准确把握特赦条件和基本要求，依法准确尽快做好特赦工作，不能错放一人，也不能
      漏赦一人。
    </SampleContent>
  </Sample>
  <Sample>
  </Sample>
  <Sample>
  </Sample>
</Samples>
```

新闻测试集

2. 输出文件格式

每组输出结果为一个扩展名为 txt 的文本文件，结果文件名称由参赛者自定，只需在 config 配置文件（config 配置文件请参见《stokis for text analysis 使用说明文档》）中注明。结果文件中的每一行对应一篇新闻的关键元素识别结果，形式为“新闻 ID 号, 实体, 实体类型, (段落位置, 起始位置, 结束位置)”。样例如下图所示，其中时间的表示形式为“yyyy-mm-dd”，若无年份则用 xxxx 表示，即“xxxx-12-28”。实体类型分为四类，LOC（地理位置），ORG（组织机构），PER（人物），TIME（时间）。其中符号均为英文半角符号。

```
73459, 西沙, LOC, (1, 12, 13)
73459, 解放军, ORG, (1, 7, 9)
73459, 选择, PER, (5, 13, 14)
73462, 南海, LOC, (1, 30, 31)
73462, 中方, ORG, (1, 42, 43)
73462, 选择, PER, (3, 13, 14)
1748, 标本兼治, LOC, (4, 202, 205)
1748, 党, ORG, (1, 5, 5)
1748, xxxx-12-28, TIME, (3, 7, 12)
```

四、评价指标

事件关键元素识别采用准确率、召回率以及 F 值作为评价指标。参赛系统的输出结果的实体集合记为 $S = \{s_1, s_2 \dots s_m\}$, 标准结果的(Gold Standard)的实体集合记为 $K = \{k_1, k_2 \dots k_n\}$, 我们定义 S 中的元素 s_i 与 K 中的元素 k_j 等价，当且仅当：

- 1) $s_i.entity = k_j.entity$ (实体)

2) $s_i.entity_type = k_j.entity_type$ (实体类型)

3) $s_i.doc_id = k_j.doc_id$ (实体对应的样本)

4) $s_i.pos = k_j.pos$ (实体的起止位置)

基于以上等价关系，我们定义集合 S 与 K 的交集为 $S \cap_{eq} K$

$$Precision = \frac{|S \cap_{eq} K|}{|S|}$$

$$Recall = \frac{|S \cap_{eq} K|}{|K|}$$

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

最终排名以综合评分的 F 值作为依据。

五、测试步骤

1、在调试阶段，参赛方远程登录客户机，进行系统部署与环境搭建，并获取调试数据进行调试。

2、在系统测评阶段，参赛方提供可执行程序。其中，使用 windows 虚拟机的队伍训练程序和测试程序分别提供 xxx.bat 的执行文件，使用 linux 虚拟机的队伍训练程序和测试程序分别提供 xxx.sh 的执行文件。参赛方需要填写代理程序的 config 文件，利用代理程序 stokis 提交评测任务。比赛数据和测试步骤请参见《stokis for text analysis 使用说明文档》。

3、参赛方允许在规定时间内完成关键元素识别任务，包括 3 组新闻测试数据集。每组测试集可以提交任意次测试结果（上限 10 次），但只有第一次结果的评测成绩作为最终排名依据。如果参赛方没有提交一组测试集的运行结果，则该组测试集的得分为 0。

说明：

1、代理程序的功能。代理程序会自动完成以下操作：1) 切断客户机的访问连接，2) 下载训练数据，3) 运行训练任务产生模型文件，4) 下载测试数据，5) 运行测试任

务产生结果文件，6) 上传结果文件到服务器，7) 提交评测服务获得评测结果，8) 将整个运行结果存入数据库，9) 恢复客户机访问连接。

2、技术指标评分。代理程序自动运行事件关键元素识别评测工具，给出相关性能指标，测试结果形式如下：

新闻第 1 组结果，准确率___，召回率___, F 值___；

新闻第 2 组结果，准确率___，召回率___, F 值___；

新闻第 3 组结果，准确率___，召回率___, F 值___；

最终评测结果按照 1:3:6 的比例对三次测试结果的 F 值进行加权平均。

六、注意事项

1、参赛方需要在指定时间内完成全部 3 组测试集的运行测试。每组测试集可以提交多次测试任务(上限 10 次),但只有第 1 次提交的测试任务的评测成绩作为排名依据。如果某一组测试集没有对应的输出结果，则该组测试集的评测成绩为 0。该项目最终得分为 3 组评测成绩的加权平均。

2、代理程序开始运行后，如果参赛程序中断，则视为一次失败的提交。参评程序执行过程中，应记录程序中断提示信息。如果代理程序出现问题，请与主办方工作人员及时沟通解决。

3、代理程序开始运行后，参赛方与客户机断开连接，无法通过调试程序来优化结果。代理程序运行完毕后，参赛方可重新登录客户机，通过调试程序来优化结果。