

中文微博实体链接评测大纲

1. 评测对象

本次评测面向中文知识库的构建与扩充技术，对中文微博中的命名实体与中文知识库的链接算法进行评测。

2. 任务设置

本次评测设置了一个核心任务：中文微博中的实体链接。在这一任务中，实体链接的目标是一个利用网络百科页面中的 InfoBox 自动构建的知识库（InfoBox 实例请参见互动百科或维基百科的页面）。具体而言：给定一条微博，一个待链接的字符串，以及该字符串在这条微博中出现的位置，本任务要求首先判断该字符串是否指向了知识库中的某一个实体；若存在这样的对应实体，则将该实体在知识库中的编号输出，若不存在，则输出空置符 NIL。

提交格式为：

id: 结果序号

team-tag: 队伍结果标识

run-tag: 提交结果组号（同一队伍可以提交至多 2 组结果）

weibo-id: 微博 id

name-id: 字符串 id（同一条微博中可能存在多个待判断的字符串）

KB-id: 知识库中的实体 id，若不存在，则标记为 NIL

（不同字段直接用制表符（\t）分隔）

例如：在如下 id 号为 1 的微博中，标识符 content 所指向的为该微博的具体内容，标识符 name 则标记了待测定的字符串（同一条微博中可能存在多个待判断的字符串），标识符 startoffset 和 endoffset 标记了其对应 id 的字符串在该微博中的具体位置；在训练数据中，每个带测定的字符串还将会配备一个 kb 标识符，用来记录该字符串在知识库中所对应的实体的编号，对于那些在知识库中找不到对应实体的字符串，则标记为 NIL（kb 标识符不会出现在测试数据中）。

```
<weibo id="1">
```

```
  <content>北京时间 3 月 12 日，2013 亚冠联赛小组赛第二轮，广州恒大足球俱乐部  
  客场挑战全北现代，广州恒大首发已经公布。</content>
```

```
  <name id="1">广州恒大足球俱乐部</name>
```

```
    <startoffset id="1">25</startoffset>
```

```
    <endoffset id="1">34</endoffset>
```

```
    <kb id="1">KBxxxx<kb>
```

```
  <name id="2">全北现代</name>
```

```
    <startoffset id="2">38</startoffset>
```

```
    <endoffset id="2">42</endoffset>
```

```
    <kb id="2">KByyyy<kb>
```

```
  <name id="3">广州恒大</name>
```

<startoffset id="3">43</startoffset>
<endoffset id="3">47</endoffset>
<kb id="3">KBxxxx<kb>
</weibo>

注：文件采用 unicode (utf-16) 编码，每个字符占两个字节，微博内容中第一个字符的 startoffset 为 0，endoffset 为 1，以此类推。比如：上述微博 (weibo1) 中的“亚冠联赛”这个字符串所对应的 startoffset 为 14，endoffset 为 18。

在上面的例子中，“广州恒大足球俱乐部”和“广州恒大”指向知识库中的同一实体“广州恒大足球俱乐部”（知识库编号 KBxxxx），而“全北现代”则指向另一个实体“全北现代汽车足球俱乐部”（知识库编号 KByyyy）；则应提交的结果为：

1	team_name	1	1	1	KBxxxx
2	team_name	1	1	2	KByyyy
3	team_name	1	1	3	Kbxxxx

3. 数据说明

本次评测数据全集包括 10 个话题，每个话题采集大约 1000 条微博，共约 10000 条微博，平均每条微博包含 1 至 2 个待测定字符串。所有数据均采用 xml 格式组织，unicode (utf-16) 编码存储。本次测试所使用的中文知识库，将会在数据样例发布时一并提供。

4. 评价标准

本任务使用微平均准确率 (micro-averaged accuracy across name strings) 来评价各队伍对于这一子任务的结果。例如，假设评测中的所有测试数据为 3 条微博中的 5 个字符串，某支队伍的系统输出结果如下：

测试数据编号	系统输出结果	人工标注结果	统计
wb1_n1	KB1000	KB1000	正确
wb2_n1	KB1222	KB1222	正确
wb2_n2	KB2000	KB2000	正确
wb3_n1	KB3111	KB3111	正确
wb3_n2	NIL	KB3111	错误

则该队伍的成绩为：4/5=0.8 。

本次评测为离线评测，参赛单位自行处理测试数据，按照约定格式提交结果。测试数据的标准答案采用人工标注的方法确定。参赛单位需要处理全部测试数据，但用于实际评测的人工标注数据约占评测数据全集的 10%左右。每家参赛单位允许最多提交 2 组结果。