

# ICTLAP\_CORE\_NE 与 Drawline 测试方案报告

林基远

June 18, 2013

## Contents

|     |   |   |
|-----|---|---|
| 1   | 模块介绍                                    | 1 |
| 2   | 使用及结果说明                                 | 1 |
| 2.1 | 使用环境                                    | 1 |
| 2.2 | 使用步骤                                    | 1 |
| 2.3 | 结果说明                                    | 1 |
| 3   | 测试方案                                    | 2 |
| 3.1 | 测试 <i>ictlap_core_ne</i> 是否正确识别出实体      | 2 |
| 3.2 | 合并实体实例到关系配置模板, 并使用 <i>drawline</i> 抽取关系 | 2 |

## 1 模块介绍

这两个模块搭配的功能是对一篇文章, 抽取其中的命名实体, 然后对这些实体按照配置的模板进行关系抽取。其中

- *ICTLAP\_CORE\_NE* 是对文章进行命名实体识别 (识别人名、组织名、地名), 识别出的实体将会合并到关系模板
- *Drawline* 是对文章按照合并的实体、关系模板抽取这些关系实例

## 2 使用及结果说明

### 2.1 使用环境

Linux-64

### 2.2 使用步骤

```
$ cp template.cfg combine.cfg
$ ./ne_driver article.txt | grep CONCEPT >> combine.cfg
$ ./drawline combine.cfg article.txt > res.txt 2>/dev/null
```

| 文件           | 说明         |
|--------------|------------|
| article.txt  | 需要抽取的文章    |
| template.cfg | 关系模板       |
| combine.cfg  | 关系与实体合并的模板 |
| res.txt      | 抽取出的关系实例结果 |

## 2.3 结果说明

默认结果格式如下。在\*\*\*\*\*上的是抽出的关系，\*\*\*\*\*下的是概念。对于每个关系实例，第一行为offset和len，即关系实例对应字符串在原文本中的byte-offset，byte-len。接下来是关系实例，冒号前是规则的参数名，冒号后是实体实例。接着以一行5个-结束该关系。

```
34440 21
action:建立
person:薛建军
-----
*****
NAME:吕尧臣
NAME:徐汉棠
```

## 3 测试方案

### 3.1 测试ictlap\_core\_ne 是否正确识别出实体

如下，article.txt是一篇文章，ictlap\_core\_ne 已识别出里面主要的人名、机构名、地名。这里也有错误的地方，如将奥巴马识别成地名，以及没有识别出热比娅这个人名。由于ictlap\_core\_ne 可能会输出其他信息，这里用grep过滤。

```
$ cat article.txt
玛依努尔接受了CNN的采访。热比娅的丈夫在接受采访的时候说，热比娅的儿子呼吁美国议会。
热比娅受到了日本总理的邀请。奥巴马约见了热比娅。热比娅说要加强与藏青会的合作。
$ ./ne_driver article | grep CONCEPT > ne.txt
$ cat ne.txt
CONCEPT:NAME:玛依努尔
CONCEPT:LOCATION:美国
CONCEPT:ORGANIZATION:美国议会
CONCEPT:LOCATION:日本
CONCEPT:LOCATION:奥巴马
```

### 3.2 合并实体实例到关系配置模板，并使用drawline 抽取关系

```
$ cat template_cn.cfg ne.txt > combine.cfg
$ ./drawline template_cn.cfg article.txt > res.txt 2>/dev/null
$ cat res.txt
164 15
person:奥巴马
action:约见
-----
1 18
person:玛依努尔
action:接受
-----
*****
NAME:玛依努尔
LOCATION:美国
ORGANIZATION:美国议会
LOCATION:日本
LOCATION:奥巴马
```

按照已有的实体及关系，这里已正确识别出。但也有的地方，比如“日本总理的邀请”没有识别出来，这个可以通过手动为combine.cfg添加实体及关系解决。