



## Rules 编写规范

(本文档会随时更新版本以时间为准，时间最新为最新版)

2018. 0428-v1.0

### 一、语义

#### 1. 句子结构

在编辑器测试单元中，有两个作用：一是测试一句话的分词情况，二是测试是否推理成功。

#### 2. 语义提取

语义提取就是提取句子中有含义的数据，比如：地名、人名、行为、判断等。

比如例句：小红想吃西瓜。这里边有两个对象，一个是主语小红，另一个是宾语西瓜，行为是吃，要明确句子表达含义。

分词后：小红[1]想[2]吃[3]西瓜[4]，可在小红和西瓜之间建立一个语义关系，定义成#[1],rule-eat-fruit,[4]。rule的写法不在此处过多介绍。

### 二、Rules 编写原则及注意事项 (加“\*”的为重点项)

1. \*语义关系的名字必须能够覆盖所有同类的关系，一个语义关系，一般要对应 100 条精简优化后的 rule,或者写更多的 rule。

2. 一般简单常用语句是不需要写 rule 的，如：吃海鲜。这种词句更适合用关键词

3. \*每一条 rule 所选取的变量在树状图里面最好是同一条线上。

4. \*每一条 rule 的变量总数最好控制在 5 个以内，其他句子成分在不影响提取要素的情况下可以[NA]掉的尽量[NA]掉，或者不做标识；所选取的提取要素以两到三个为最佳，过多影响引擎语义推理速度，过少会增加无用 rule 条数。

5. 语义关系不要定义太多，不能每一句话都定义语义关系，含义基本一致的，要用一个语义关系表达（有几个字差别的可以在关键词单元做映射或者强制转换）。

6. \*值域不能限制的太死，更不能只用当前句子中的词来定义值域，极少数特殊情况时可以考虑直接限定死或者考虑关键词解决。
7. \*带有“什么”“哪些”等含有疑问意图的句子，写 rule 时，要写两条 rules，并且其中一条 rule 前缀一定要加“QA:”。
8. \* rule 有冲突时,可以通过 test 测试单元先了解是和哪些 rule 有冲突,可以考虑增加限定解决，对值域或者关键词做调整。
9. \* 除了语义关系两端的节点，另外的词都用同类词限定值域，词性不相同的词不能放在同一个值域中；如果一定要放在同一个值域中，有可能会令这条 rule 无效。但是想这样使用，可以考虑在关键词单元定义词性。或者把不同词性的值域分开写 rule（转换词性可能影响其他 rule 使用，不建议转换词性，建议词性不同的分开写）
10. 语义关系两端的节点尽量不限制值域，或者值域要比较广泛。#[1],rule-eat-fruit,[4]，就是说[1]和[4]两个变量最好不限制值域，即便限定也要设置很宽。如：4=++水果，水果的范围很广。
11. 一个规则，除了特殊需要，条件模板一般都在 3—5 个节点，少于 3 个，则很容易错误识别，大于 5 个则通用性就比较差了，这还是值域设定比较合理的前提下。
12. \*项目中写 rule 的时候，首先我们要清楚对话场景是什么，我们要提取的语义要素是什么，根据语义要素制定节点，选取变量。
13. \*一句话里边有很多个语义要素，我们选取的是必须提取的意义要素作为变量。  
比如例句：我晚上要去图书馆。这里边就有两个要素是我们必须提取的：1）时间（晚上），2）图书馆（地点），至于对象（我）可以设置为变量，但不做提取。  
如：我晚上要去图书馆  
1) 我[1]晚上[2]要去[3]图书馆[4]2=++时间,4=++地点~#[2],rule-find-go[4].  
这种是同时提取两个变量。写 action 时可以对应两个返回参数。  
2) 我[1]晚上[2]要去[3]图书馆[4]2=++时间,4=++地点~#visitorname,rule-find-shijian[2].  
我[1]晚上[2]要去[3]图书馆[4]2=++时间,4=++地点~#visitorname,rule-find-didian[4].  
这种是单一变量提取（其中一个变量为固定#visitorname），同一句话写多条 rule，一个变量对应一条 rule。  
两种写法看情况而定。
14. \*测试 rule 时，首先原句匹配测试成功，然后同样语义换句子表达测试，通过后进行长句多要素匹配测试。

### 三、rules 编写具体编写示例（只写需特殊处理的例子）

**说明：**在写 rule 的时候，经常会出现 rule 的书写格式是正确的，但是就是**匹配不上**，有的甚至原剧都匹配不上。对应以上原则和注意事项出现这种情况主要有以下几种原因：

- 1) **Rule 书写格式不正确情况不做表述。**
- 2) **变量选取类：**每一个 rule 所选取的变量必须是在同一条线上的，可根据树状图选取。
- 3) **值域类：**一种是定义太死，导致其他同类词无法匹配，可以扩充值域解决，测试方法是值域删掉看是否匹配，匹配则为该种类。**另一种是**所定义的值域包含与所选词性不同词性的词。找出不同词性的词，删掉或者在关键词单元定义词性。
- 4) **句子结构有变化：**在某些情况下有些同值域的词虽然此型相同，但是会引起句子结构的变化，会导致 rule 匹配无效，这样就需要针对个别词单独写 rule，或者对该词做强制转换，选取哪一种视情况而定。
- 5) **引擎分词错误类：**这种情况是由引擎分词是导致的匹配不成功，这种一般情况下需要对引擎拆词错误的词进行词性定义，词性定义无效情况下则需要对该词进行强制转换。
- 6) **需要写 rule 的句子分词时有断点：**这种情况一般只需要将断裂处两端的词定义成一个词即可。

**说明：**编辑器其他具体实例和其他问题详见“编辑器问题搜集与整理.docx”文档。

## 具体实例如下：

### 1. 引擎分词错误情况

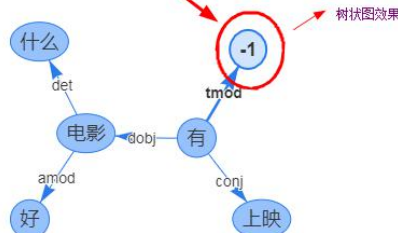
如下图：“下礼拜有什么好电影上映”这句中引擎分词错误，“下礼拜”未显示在树状图中，这就可能导致我们所写的 rule 匹配不上的情况。解决方法有两种：一种是将“下礼拜”做词性定义，另一种是将“下礼拜”强制转换成引擎可正确分词的近义词，如：下星期，这样就可以只写“下星期有什么好电影上映”这句话的 rule。类似的近义词还有很多，如：下周、下个星期等等，学习者可自行测试练习，自行体会。我所做处理只做示例不完全附在下面。

类似的引擎分词错误的还有很多，如：最近。都可以做以上处理。

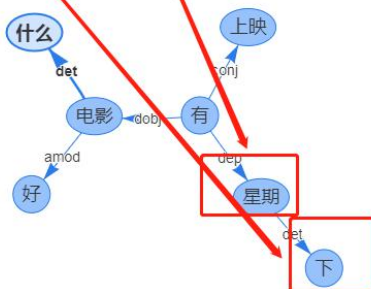
在关键词单元写入：“=下礼拜|副词”或者“=\*\*下礼拜|下星期”

Sen: 1: - 下礼拜有什么好电影上映  
RELS: tmod (有-2-VV, -1-NT) det (电影-5-NN, 什么-3-DT) amod (电影-5-NN, 好-4-JJ) dobj (有-2 -VV, 电影-5-NN) conj (有-2-VV, 上映-6-VV)

分词效果



Sen: 1: - 下星期有什么好电影上映  
 RELs: det (星期2-NN, 下-1 -DT) dep (有-3-VV, 星期-2-NN) det (电影-6-NN, 什么-4-DT) amod (电影-6- NN, 好-5 -JJ) dobj (有-3-VV, 电影-6-NN) conj (有-3-VV, 上映-7-VV)



## 2. 引擎分词导致句子断开

这种情况处理起来比较容易，至于要将“选择”定义词性即可。在关键词单元输入“=选择|动词”即可。

例子继续补充中……



TAURS

人 马 互 动