1. 事件

是促使事情状态和关系改变的条件。

1. 事件抽取的任务可以分为2大类：
   1. 事件识别和抽取：input描述事件信息的文本，output时间、地点、人物、与之相关的动作/状态的改变。
   2. 事件检测和追踪：将文本分解为：事件、新（不可预见）的事件。  
      分为：从按照时间排序的新闻文档中发现以前没有识别的事件；从实时新闻流中实时发现新的事件。
2. 核心概念  
   Event Mention:描述事件的词组/句子/句群

Event Trigger:一般是动词/名词，Event Mention中最能代表事件发生的词汇

Event Argument:事件的重要信息/entity mention,主要由实体、属性值等表达完整语义的细粒度单位组成

Argument Role: Event Argument在事件中扮演的角色，Event Argument与Event的关系

Event Type

1. 事件抽取  
   从文本中找到信息，进行填表

EX:

Text: 苹果公司将于西部时间9月12日上午10点，北京时间9月13日凌晨1点举行新品发布会。这一次的发布会地点是全新建造的史蒂夫·乔布斯剧院。根据目前的消息，这次发布会上苹果发布iPhone8(命名不确定)，iPhone7, iPhone 7sPlus, Apple Watch以及全新的AppleTV.

Table:

|  |  |
| --- | --- |
| 事件类型 | 发布会 |
| 公司 | 苹果公司 |
| 地点 | 西部时间9月12日上午10点 |
| 地点 | 史蒂夫·乔布斯剧院 |
| 产品 | iPhone8, iPhone7s. iPhone7s Plus, Apple watch3, apple TV |

1. 基于模式匹配的方法

最开始，基于人工编写的规则、语法树、正则表达式

Ex: CIRCUS. RAPIER, SRV.AUTOSLOG, LIEP.PALKA,CRYSTAL,HASTEN

后来，有监督的学习模型，ACE，但是由于标注一致性的问题，系统的效果普遍较差

基于模板的抽取方法：通过句法syntactic和语义约束 semantic constraints来识别

1. 基于人工标注语料

从一个大的标注集开始创建模板。模板的产生完全基于人工标注语料。

1. AutoSlog(Riloff)

基本假设：①事件首次提及之处即可确定该元素与事件间的关系。②事件周围的语句中包含了事件元素在事件中的角色描述。

通过监督学习和人工审查来建立抽取规则。通过训练数据中已经填充好的槽，来解析槽附近的句法结构，从而自动生成抽取规则。这个过程中抽取产生的模板过于general,所以需要人工审核。

本质上形成的是一个字典。

1. PALKA

基本假设：特定领域中高频出现的语言表达方式是可数的。

用语义框架和短语模式结构来表示特定领域中的抽取模式。通过融入WordNet的语义信息，PALKA在特定领域可以取得接近纯人工抽取的效果。