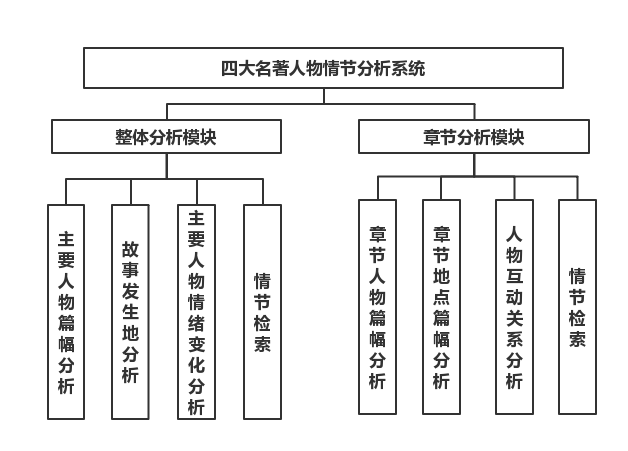
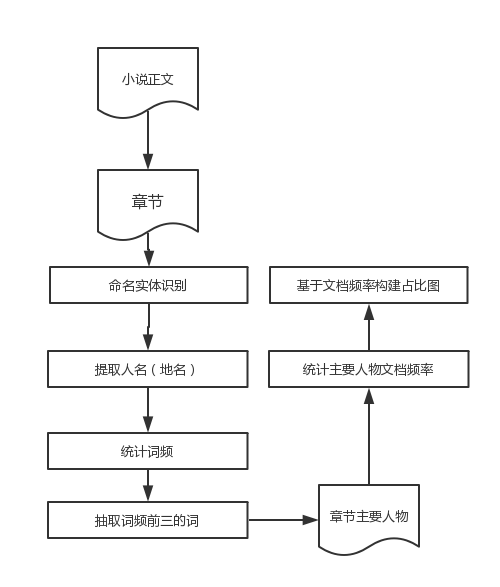
**pyltp-books**

人物篇幅分析，故事发生地分析，人物互动情况分析 [](https://github.com/gugug/pyltp-books/blob/master/PrtScr/system-structure.png)

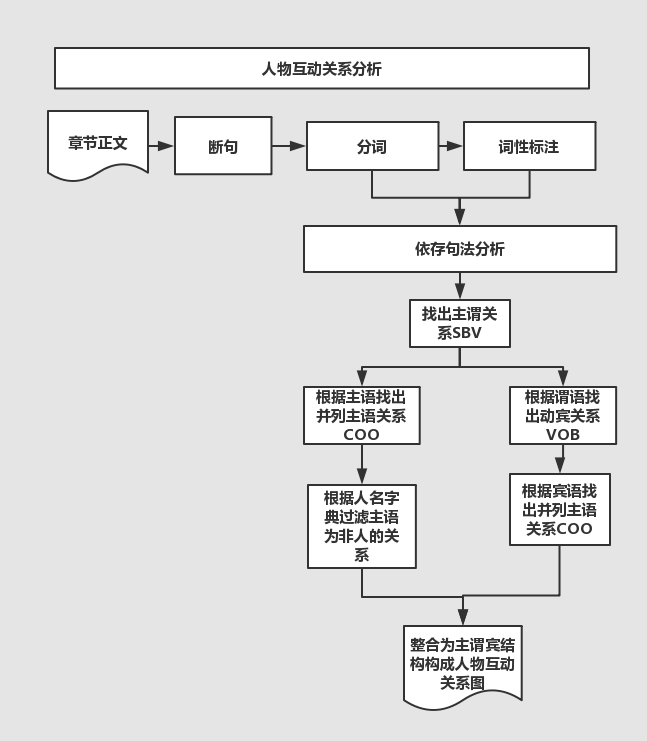
**主要人物（地点）篇幅统计策略**

本文对小说主要人物（地点）的衡量指标有两个，第一个是在文中出现的次数（词频），第二这是在文中的覆盖范围（文档频率）。基于这二者，篇幅占比的统计策略如下：  
[](https://github.com/gugug/pyltp-books/blob/master/PrtScr/main-peo-local.png)   
步骤：

1. 提取小说每一章节，进行命名实体识别
2. 抽取其中的人名（地名），并统计每个章节人名（地名）频率
3. 抽取词频最高的前三个人名（地名）作为章节主要人物
4. 对所有章节的主要人物进行文档频率统计（章节数）
5. 基于文档频率，构建篇幅占比图

**人物互动关系识别策略 （这边其实就干了提取**

**主谓宾 结构（其中主语是人物）这么一件事）**

基于依存句法树进行人物互动关系的识别，主要提取特定句子成分（主谓宾），策略如下： [](https://github.com/gugug/pyltp-books/blob/master/PrtScr/relation-action.png)

这边弄个COO把代码搞得很复杂，其实意思就是 本来有一个 主谓宾 结构

小张和小王 喜欢 小花 这边 小张和小王 就是一个COO，如果代码里面不加COO的话，我们找出来的 主谓宾 结构只有 小王 喜欢 小花

步骤：

1. 对章节正文进行断句
2. 对句子进行分词，词性标注以及依存句法分析
3. 基于SBV，找出主谓关系

|  |  |  |  |
| --- | --- | --- | --- |
| 主谓关系 | SBV | subject-verb | 我送她一束花 (我 <-- 送) |

1. 基于VOB， 找出动宾关系

|  |  |  |  |
| --- | --- | --- | --- |
| 动宾关系 | VOB | 直接宾语，verb-object | 我送她一束花 (送 --> 花) |

1. 基于COO，找出主语和宾语的并列词 (我感觉)

|  |  |  |  |
| --- | --- | --- | --- |
| 并列关系 | COO | coordinate | 大山和大海 (大山 --> 大海) |

1. 将对应的主谓关系和动宾关系整合为主谓宾结构，并构造互动关系图

https://github.com/gugug/pyltp-books