Pipeline：

输入-剧本

预处理

词频分析-中文分词-词库

情感分析

主题发现

指数设计

输出-可视化呈现

Task for 1st week：

1、文本挖掘，jieba

2、文本挖掘，机器学习，SKLearn中的学习器

Knowledge：

[1] 典型的文本挖掘方法包括文本分类，文本聚类，概念/实体挖掘，生产精确分类，观点分析，文档摘要和实体关系模型（即，学习已命名实体之间的关系）。

references：

[1] 用 Python 做文本挖掘的流程,https://zhuanlan.zhihu.com/p/19630762

[2] 中文分词jieba, <https://github.com/fxsjy/jieba>

[3] SKLearn, <http://scikit-learn.org/stable/supervised_learning.html>

Tools:

[1] GitHub, <https://github.com/>

[2] GitHub for Windows, <https://desktop.github.com/>