**基于深度学习的自然语言处理**

**第3次实验报告**

**一、实验名称**

使用并修改开源工具完成依存分析。

**二、实验日期**

2020.10.24 – 2020.11.06

**三、实验目的**

（1）熟悉依存分析

（2）熟悉开源工具的使用

**四、实验数据**

提供实验的新闻数据

**五、实验步骤及结果分析**

5.1 进入百度网盘下载语料content.txt

链接:https://pan.baidu.com/s/1QyVnxUTrJ2Dqj7FIpzuhgg 密码:lekc

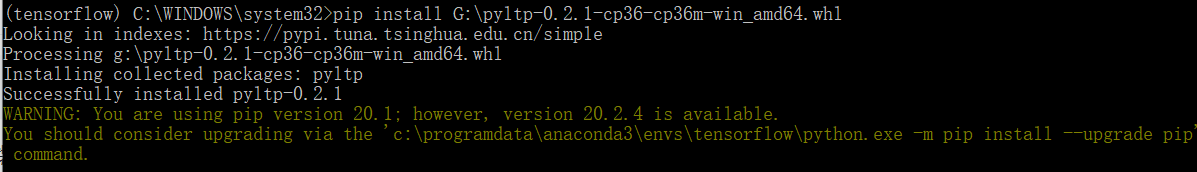
5.2 对语料进行预处理：去掉一些无用的标点符号

5.3 使用LTP进行句法分析

5.3.1 参考链接 https://blog.csdn.net/asialee\_bird/article/details/102610588

5.3.2 LTP基础

5.3.3 pyltp安装



5.3.4 pyltp实现句法分析

模型使用的LTP3.4.0,地址如下<https://blog.csdn.net/qq_32005671/article/details/73770536>

# （以上解决问题1：Segmentor: Model not loaded!）

**问题2：语义角色标注问题出现**

RuntimeError: incompatible native format - size of long

解决：https://blog.csdn.net/qq\_22521211/article/details/84029904

5.3.5 结果分析

样例结果如下：结果分别为：

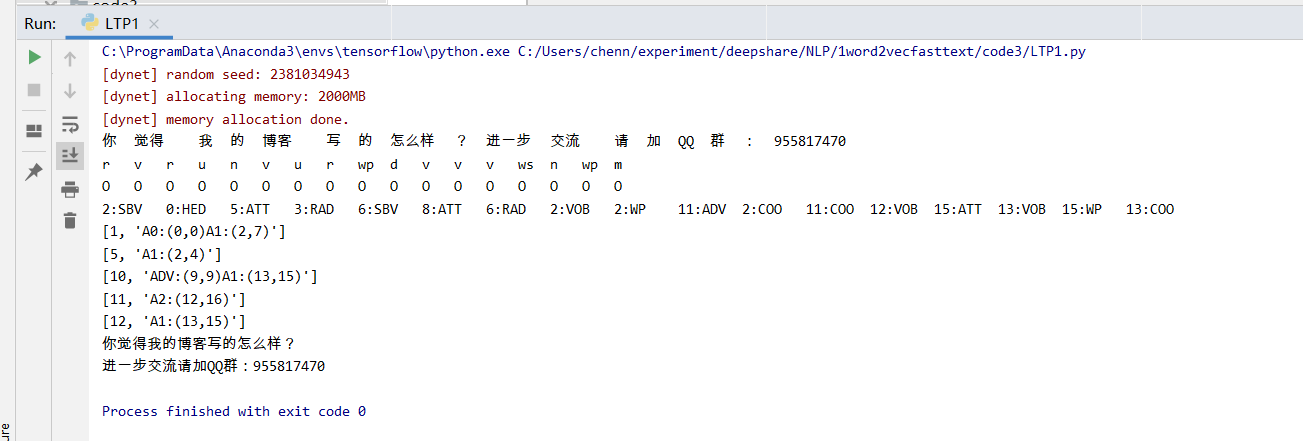
1 命名实体识别

2词性标注

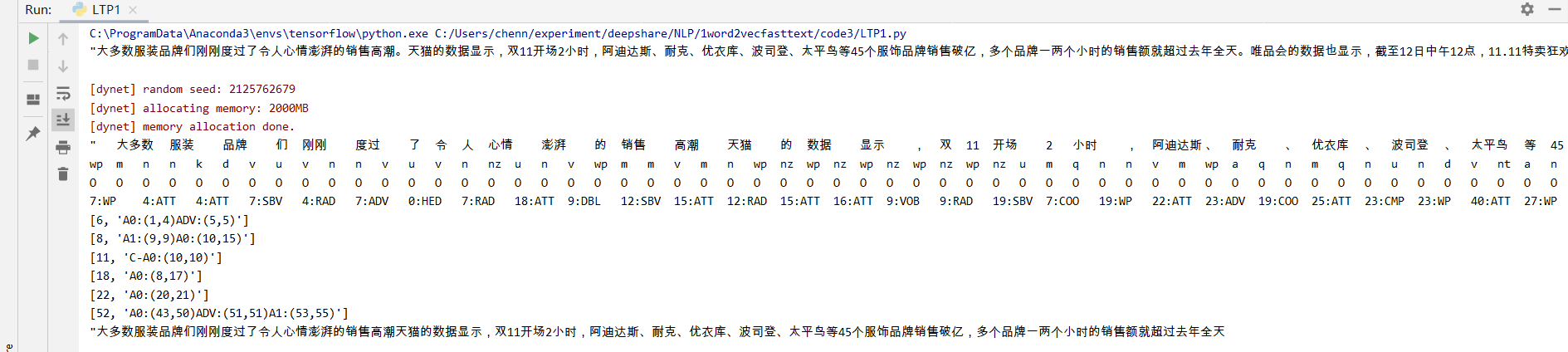
3依存句法分析

4语义角色标注

5分句



在content.txt部分数据上结果如下：



5.4 使用Stanford NLP进行句法分析

5.4.1 参考链接 https://blog.csdn.net/zkq\_1986/article/details/81583725

5.4.2 NLTK和StandfordNLP简介

**NLTK**：由宾夕法尼亚大学计算机和信息科学使用python语言实现的一种自然语言工具包，其收集的大量公开数据集、模型上提供了全面、易用的接口，涵盖了分词、词性标注(Part-Of-Speech tag, POS-tag)、命名实体识别(Named Entity Recognition, NER)、句法分析(Syntactic Parse)等各项 NLP 领域的功能。

**Stanford NLP：**由斯坦福大学的 NLP 小组开源的 Java 实现的 NLP 工具包，同样对 NLP 领域的各个问题提供了解决办法。斯坦福大学的 NLP 小组是世界知名的研究小组，能将 NLTK 和 Stanford NLP 这两个工具包结合起来使用，那对于自然语言开发者是再好不过的！在 2004 年 Steve Bird 在 NLTK 中加上了对 Stanford NLP 工具包的支持，通过调用外部的 jar 文件来使用 Stanford NLP 工具包的功能。本分析显得非常方便好用。

5.4.3 安装配置过程中注意事项

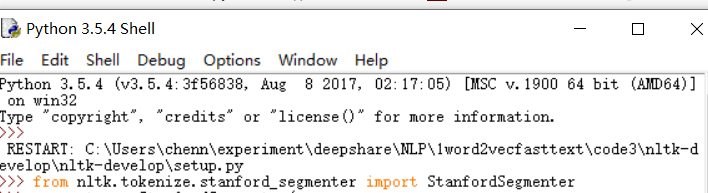
5.4.4 StandfordNLP必要工具包下载

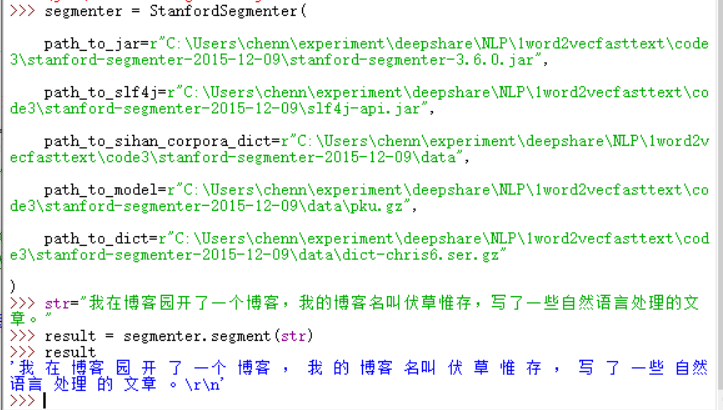
5.4.5 StandfordNLP相关核心操作

5.4.6 StanfordDependencyParser中文依存句法分析（我把所有任务都跑了一遍）

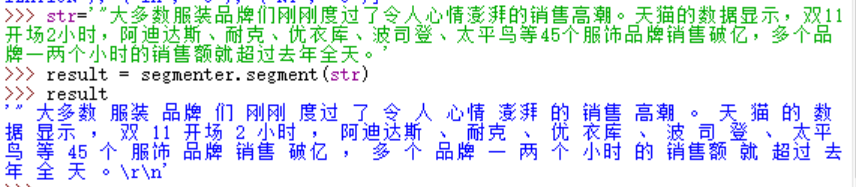
样例：

**StanfordSegmenter 中文分词**



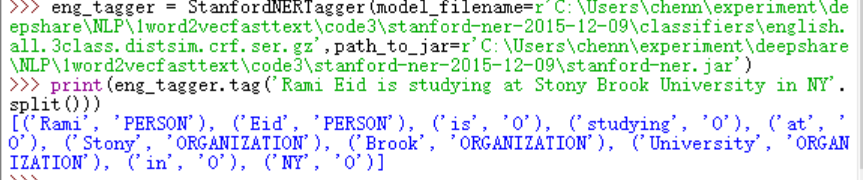


作业的数据集如下：



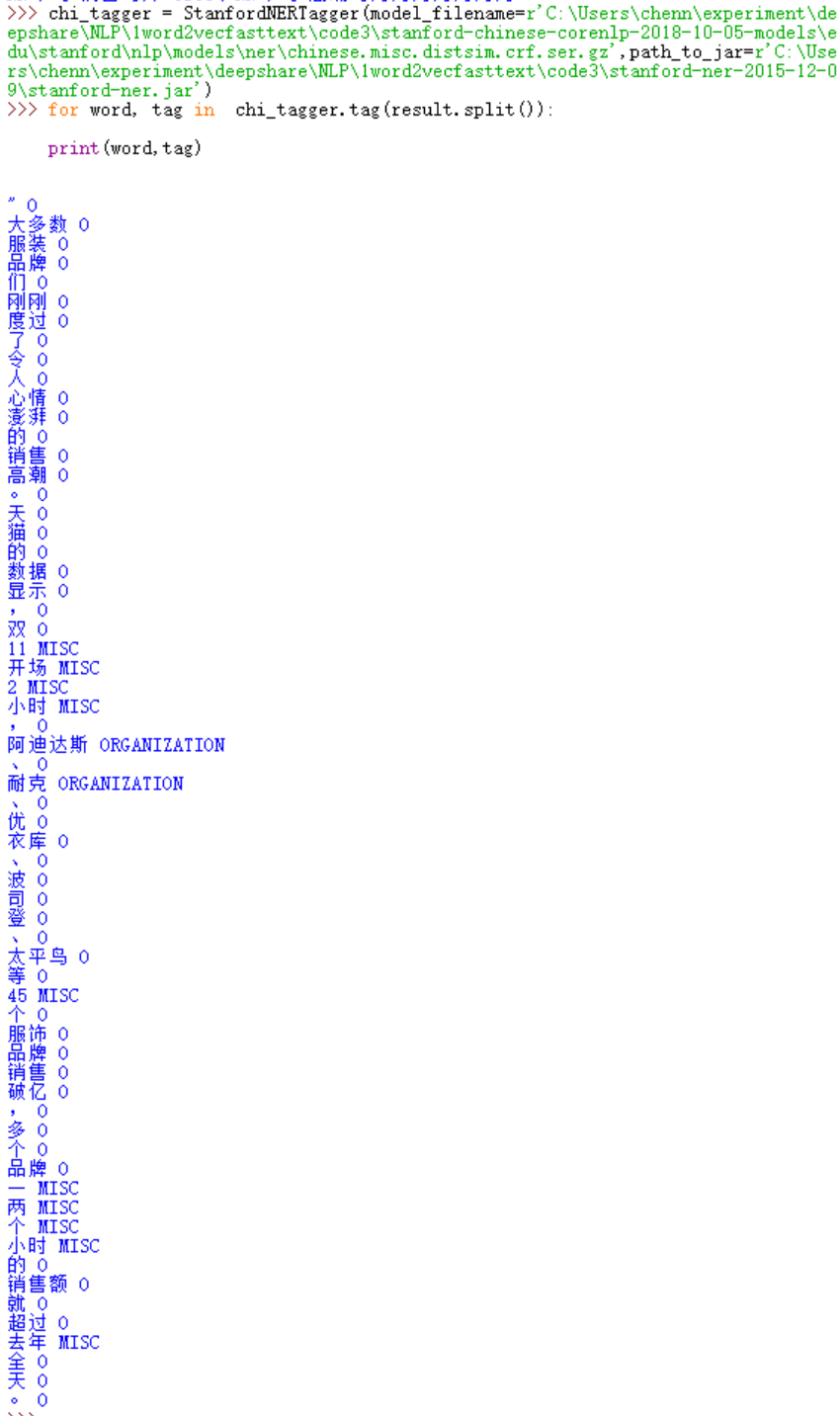
**StanfordNERTagger 英文命名实体识别**





**StanfordNERTagger 中文命名实体识别**





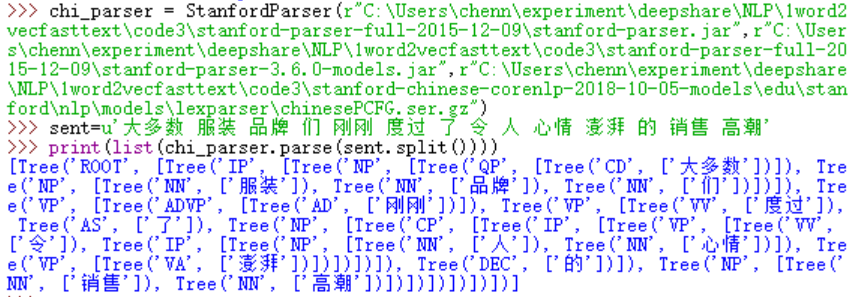
**StanfordPOSTagger 中文词性标注**





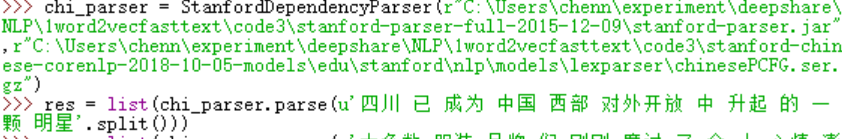
**StanfordParser 中文句法分析**

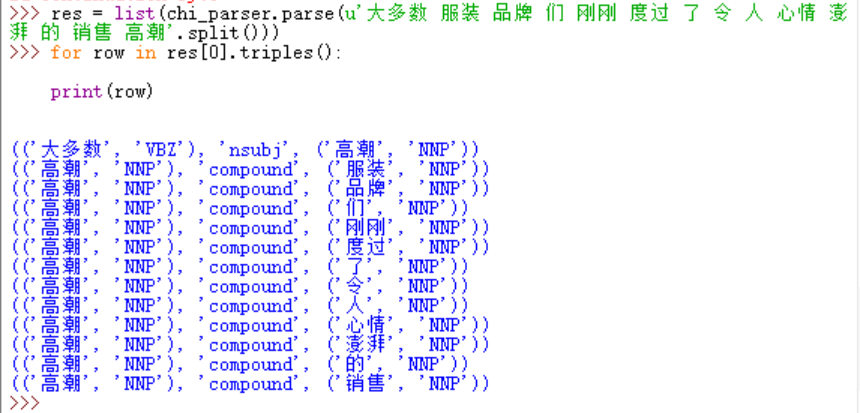


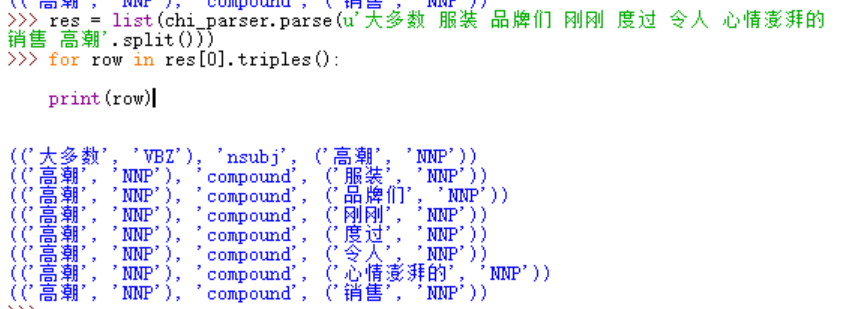


**StanfordParser中文依存句法分析**









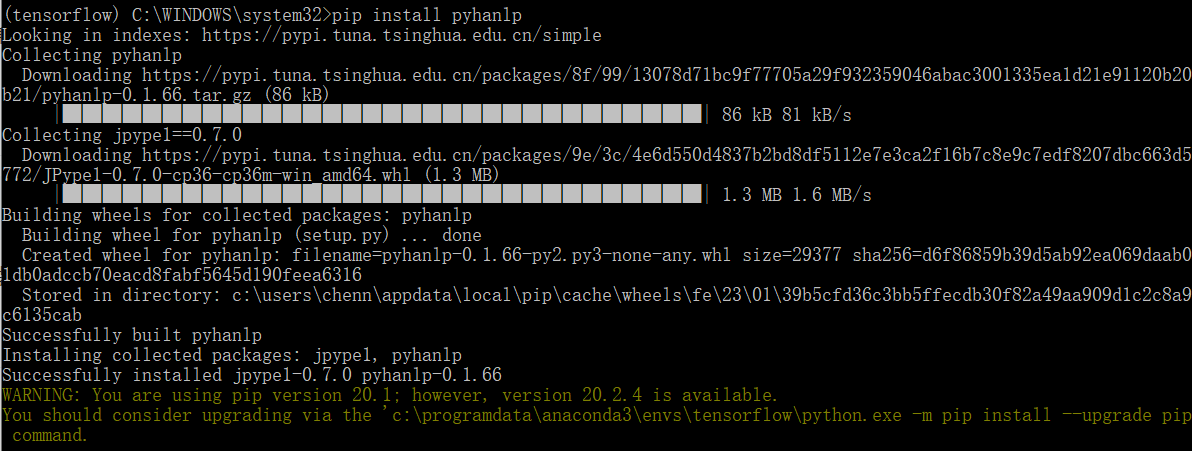
可以看到效果很差

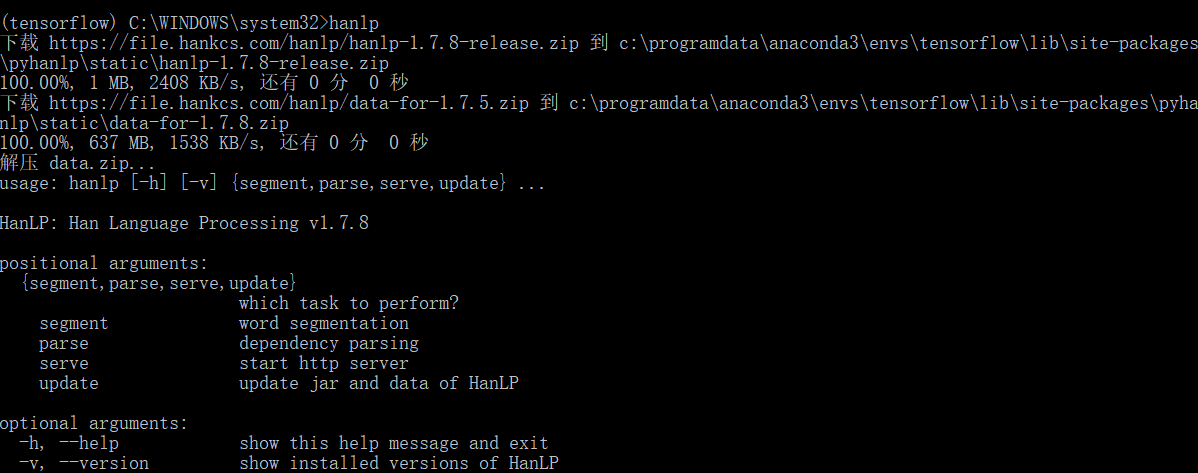
5.4.7 结果分析

5.5 使用pyhanlp进行句法分析

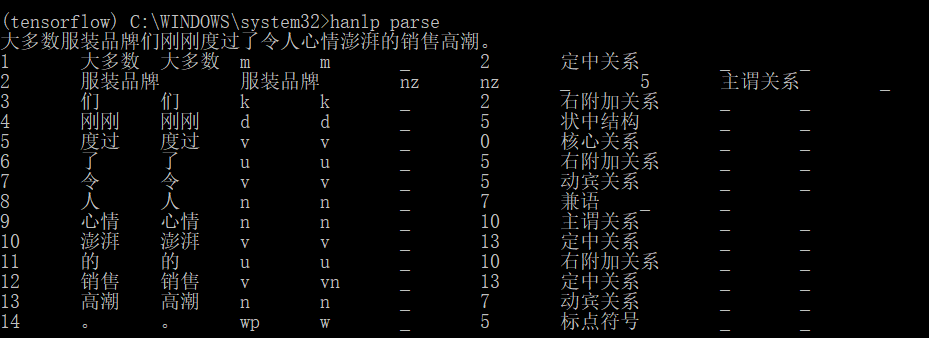
5.5.1 参考链接 https://github.com/yangby888/handnlp

5.5.2 安装





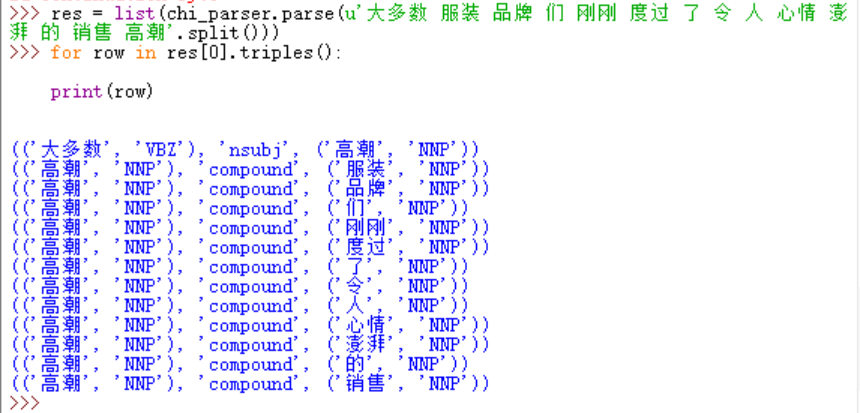
5.5.3 依存句法分析

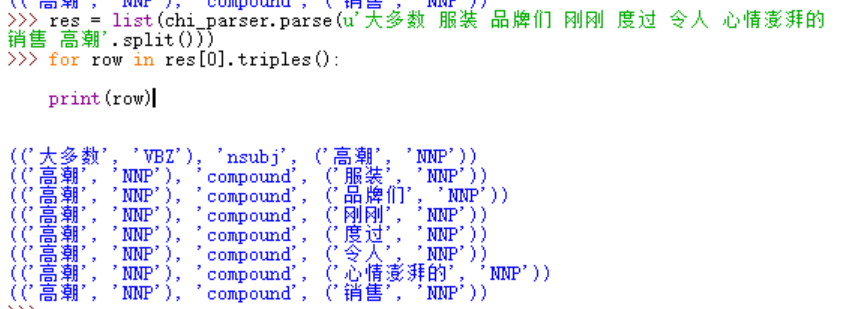


5.5.4 结果分析

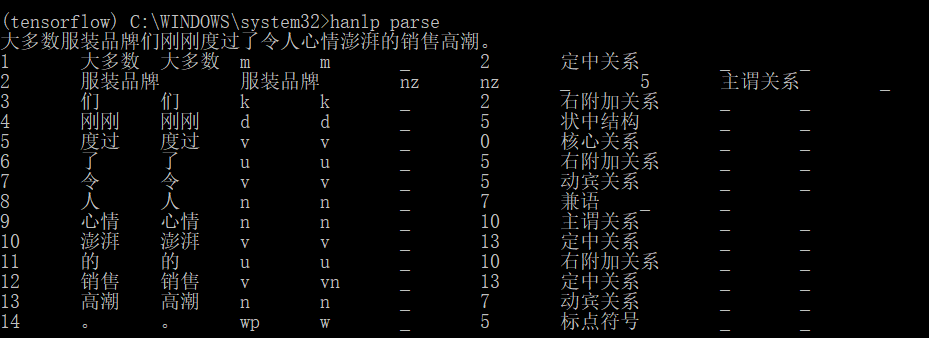
5.6 三种方法的分析与比较

**StanfordParser的依存句法分析效果如下：**

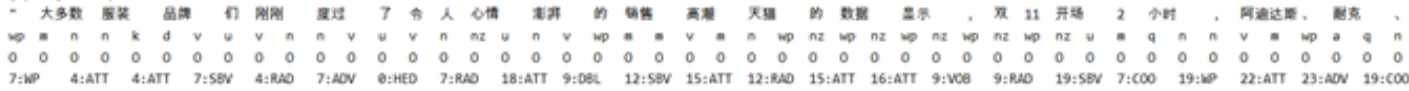




**pyhanlp的依存句法分析效果如下：**



**LTP实验效果如下：**



**可以看出StanfordParser效果差，而pyhanlp与LTP效果较好。**

**六、实验感想**

## 依存句法分析

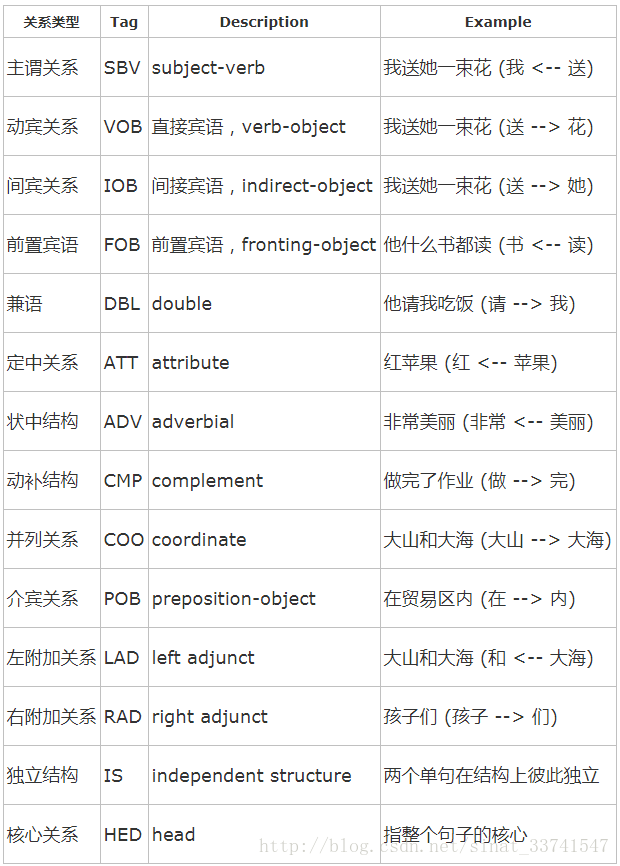
### 1、概念

依存句法通过分析语言单位内成分之前的依存关系解释其句法结构，主张句子中核心动词是支配其他成分的中心成分。而它本身却不受其他任何成分的支配，所有受支配成分都以某种关系从属于支配者。

### 2、五个条件

（1）一个句子中只有一个成分是独立的  
（2）句子的其他成分都从属于某一成分  
（3）任何一个成分都不能依存于两个或两个以上的成分  
（4）如果成分A直接从属成分B，而成分C在句子中位于A和B之间，那么，成分C或者从属于A，或者从属于B，或者从属于A和B之间的某一成分  
（5）中心成分左右两边的其他成分相互不发生关系

### 3、标注关系



**模型结果StanfordParser效果差，而pyhanlp与LTP效果较好。**