**工作汇报**

# 观测信息提取

首先，因为NSDC上的数据是以content+data的双文件形式存储的。编写了一个类似于爬虫的文本分析文件，将每一个content 的相关数据都整理到了一个文件，我称之为information。Information里包含了时间系统、坐标系统和观测台等信息，方便在某些数据记录中不包含某些数据时补上。

Information中包含了以下字段的信息

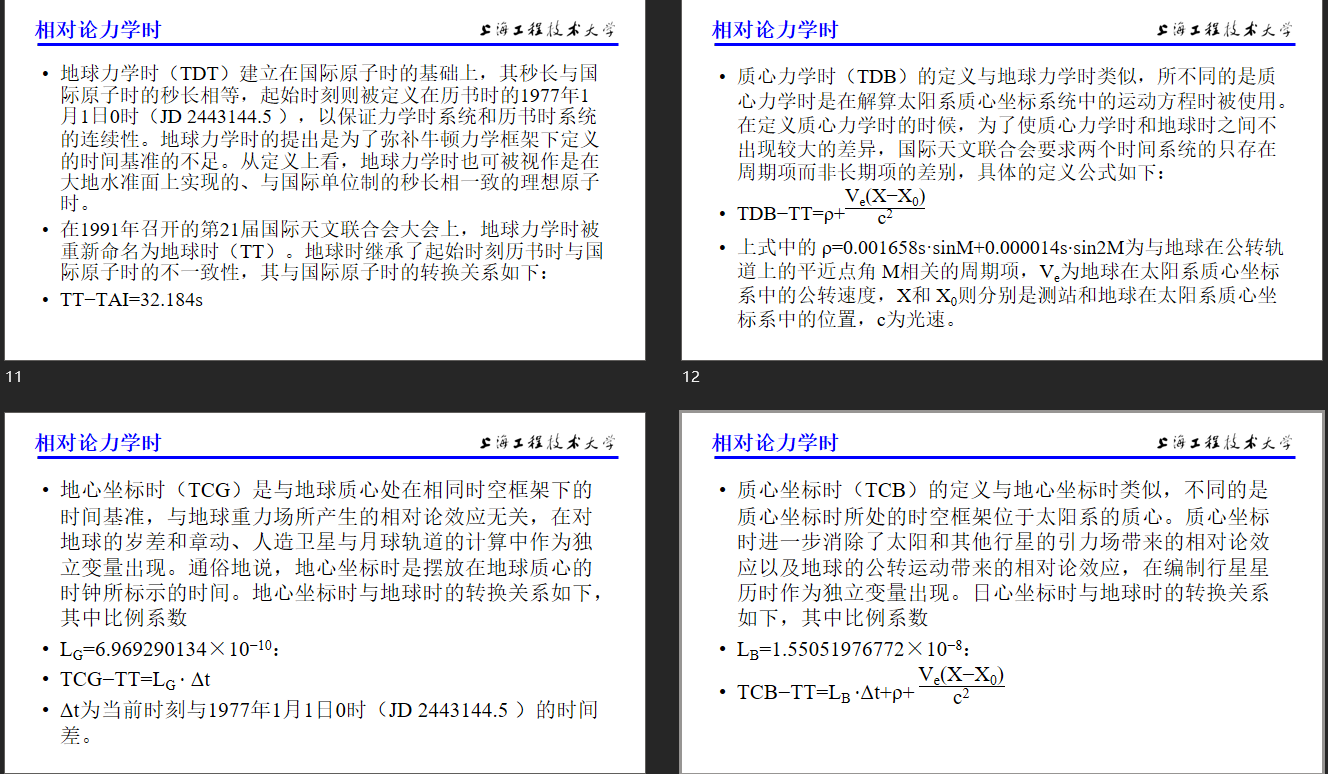
Id、Type、Dates、Observatory、Reference Frame、Centre of Frame

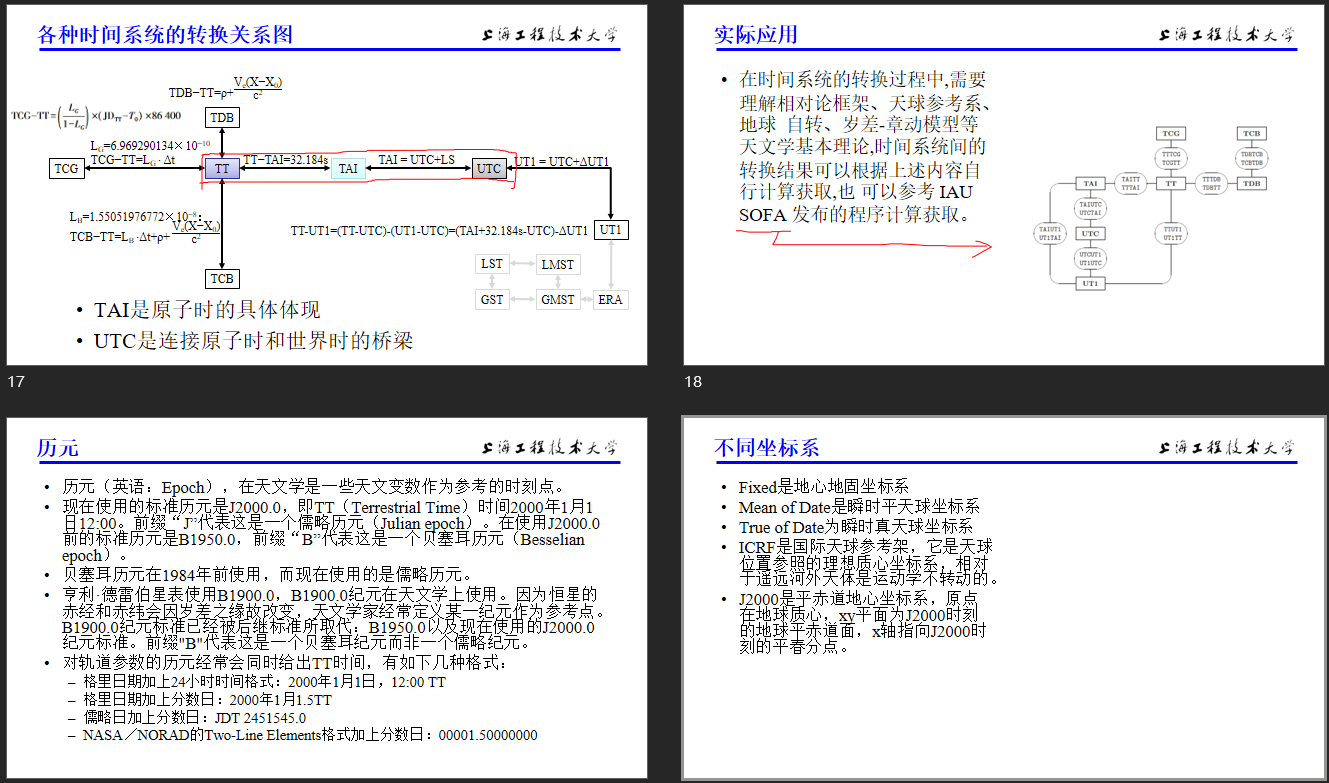
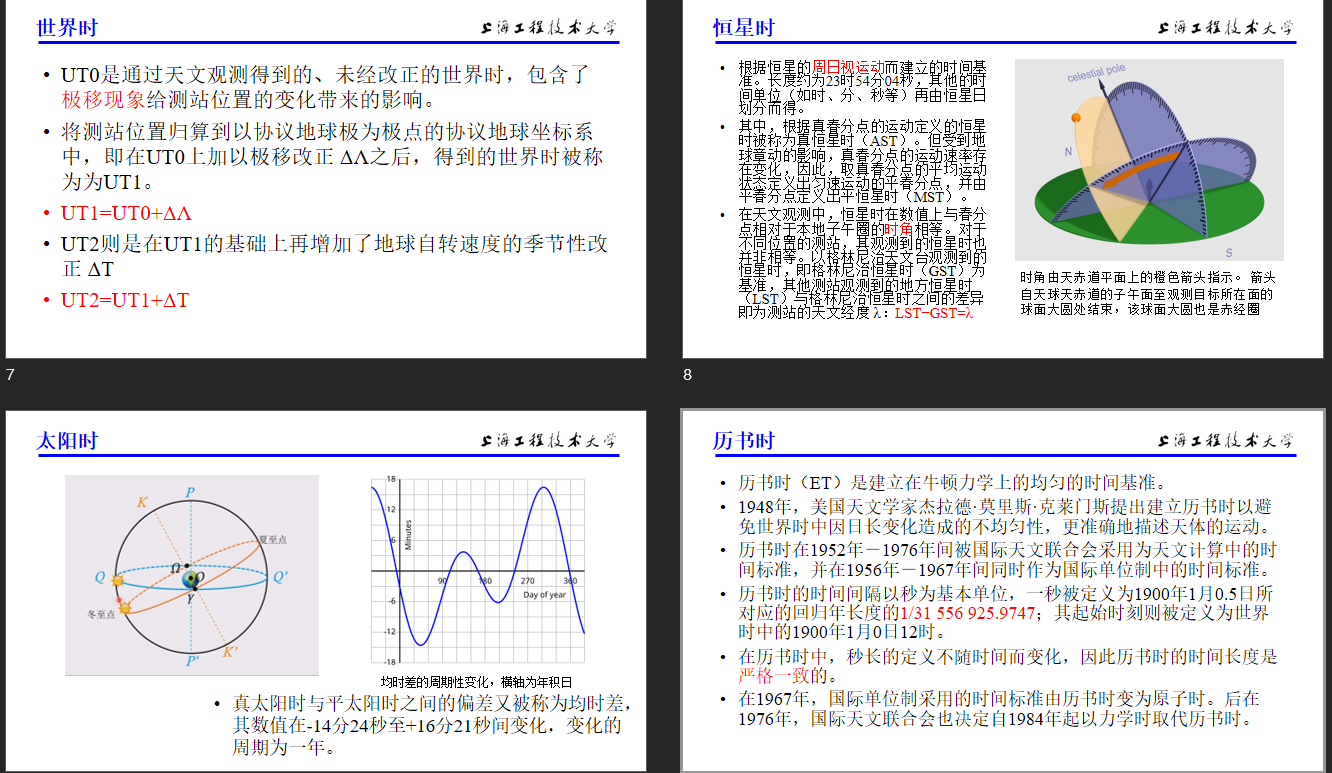
Epoch of Equinox、Time Scale、Reduction、Coordinates、Diffraction

Receptor、Telescope、Observers、Relative To



其次，因为最一开始并不了解时间系统和坐标系统的概念和定义，我统计出了information中出现的所有时间系统的类型，并制作了一个ppt在组会讲解。



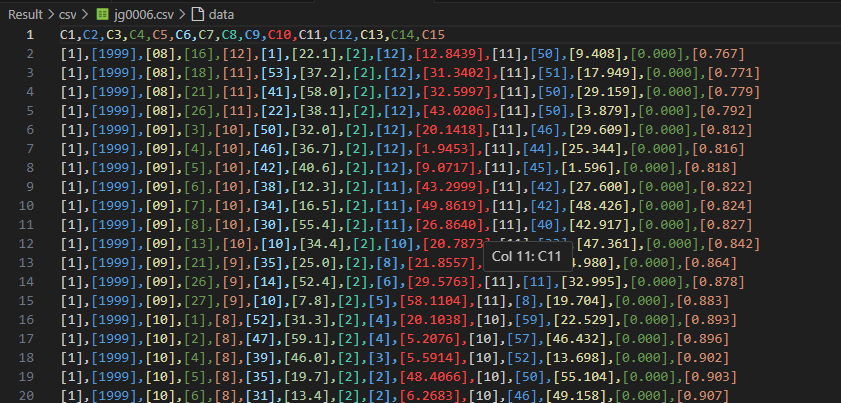


# 观测数据转换

Python对txt文件并不算特别友好，首先我打算将txt文件转换成了csv文件方便之后的各种操作，但是转换途中发现有的观测数据的格式并不完整，编写了一个csv error程序分析格式错误并改正。



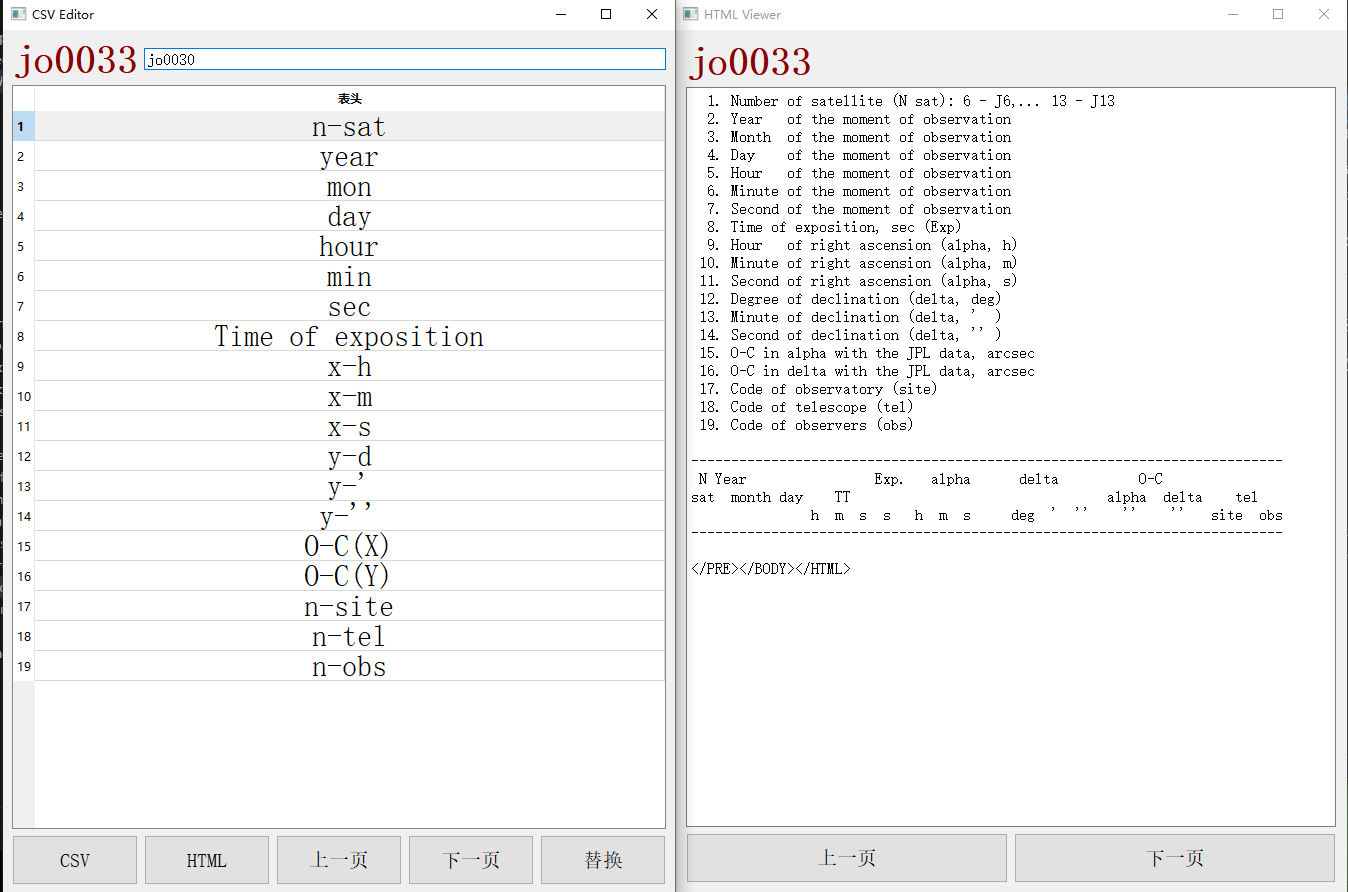
并给所有的csv文件添加了一个表头，便于根据表头实现对多表的合并。转换之后的csv文件如下，我为每个数据加上前后括号强制转换为文本，防止数据进度丢失。



# 观测数据整理

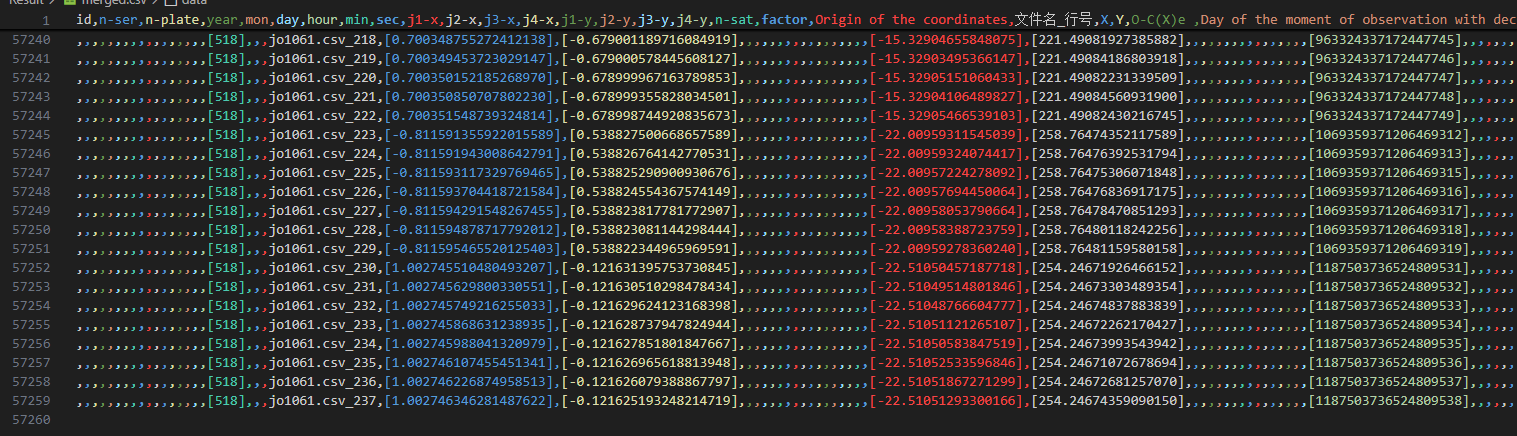
为了批量处理时间到统一的格式，亦或是方便去对不同的数据类型处理，例如有些数据是赤经赤纬，有些数据是空间坐标。还是要将Data中的数据对应上相应的列名，我编写了一个程序首先给所有csv文件增加一行标题行，然后编制出一个小软件，可以直接更改各个文件的表头，另外为了防止出错在方便同步输出html文件新，最后还加上了二次验证确保所有的数据是准确无误的。

替换表头。



# 合并观测数据

基于上述的数据准备，编写一个简单的程序就能实现多表合并，我还给所有的行都新增了一列id，以文件名+行号形式构成，方便对数据的检索和对某些指定的数据整理。整合之后的表格如下。一共包含了内部、外部等十几个卫星的5万多条观测数据。



即已经基本实现了对观测数据的整合。