

# 语言

Python, utf-8 编码

# 方法

正则表达式

# 输出格式

共一个 out.txt 文件

每一篇文书，首先包含**原文书的头部信息**；  
然后给出若干次作案的信息，**每次作案包含时间、人物、地点、盗窃物品及其价值、价值补充**，（其中的价值补充，可能是总价值，也可能是某些盗窃物品价值）；  
一篇文书中的**每次作案之间用#####分隔**；  
**每篇文书之间用\*\*\*\*\*分隔**。

例如，前六篇文书的信息抽取结果如下：

李世菊盗窃案      D9E739A9-BF89-4A58-9212-45018AAACE8D true  
accusal: 0010502\_盗窃罪,  
clause: 第六十五条, 第六十七条, 第二百六十四条,  
李世菊盗窃案  
昆明市呈贡区人民法院  
刑事判决书

（2013）呈刑初字第 40 号  
公诉机关昆明市呈贡区人民检察院。

#####

【时间】：	2012 年 10 月 17 日
【作案人】：	李世菊
【地点】：	云南省中医学院呈贡校区食堂
【盗窃物品（价值）/现金】：	李艳红的一部粉色三星牌手机,手机价值人民币 1090 元
【价值补充】：	

\*\*\*\*\*

龚伸斌盗窃案 594F3A85-C9E7-4D0B-AE0A-734F82F8667E true

accusal: 0010502\_盗窃罪,

clause: 第六十五条, 第六十七条, 第二百六十四条, 第五十二条,

龚伸斌盗窃案

浙江省杭州市拱墅区人民法院

刑事判决书

(2013)杭拱刑初字第854号

公诉机关杭州市拱墅区人民检察院。

#####

【时间】： 2013年5月29日14时许

【作案人】： 龚伸斌

【地点】： 本市拱墅区祥符街道新文村毛家里76号3楼303室

【盗窃物品(价值)/现金】： 被害人王某房间内的黑色联想笔记本一只及现金人民币500元

【价值补充】：

\*\*\*\*\*

田光军等盗窃案 31243281-482D-4C57-AB16-CE3D104A7BC2 false

accusal: 0010502\_盗窃罪,

clause: 第二十五条, 第六十七条, 第二十三条, 第二百六十四条, 第六十四条,

田光军等盗窃案

福建省南安市人民法院

刑事判决书

(2014)南刑初字第61号

公诉机关福建省南安市人民检察院。

#####

【时间】： 2013年8月中旬的一天中午

【作案人】： 田光军,赵洪军

【地点】： 位于南安市石井镇溪东村的福建南安市华星石材有限公司车间

【盗窃物品(价值)/现金】： 该公司铁块部件1个(无法估值)

【价值补充】：

#####

【时间】： 2013年8月中旬的一天中午

【作案人】： 田光军,赵洪军

【地点】： 位于南安市石井镇溪东村的南安市石井阳明石材厂车间

【盗窃物品(价值)/现金】： 该厂夹盘19个(价值人民币171元)

【价值补充】：

#####

【时间】：2013 年 8 月中旬的一天中午  
【作案人】：田光军,赵洪军  
【地点】：位于南安市石井镇溪东村的福建省泉州厦东石材有限公司车间叉车  
【盗窃物品（价值）/现金】：该公司铁饼 1 个（价值人民币 20 元）  
【价值补充】：

#####

【时间】：2013 年 8 月中旬的一天中午  
【作案人】：田光军,赵洪军  
【地点】：南安市石井镇前房村李某某经营的石材厂  
【盗窃物品（价值）/现金】：李某某暂放在该处的电机 1 个（价值人民币 1610 元）  
【价值补充】：

#####

【时间】：2013 年 8 月中旬的一天中午  
【作案人】：田光军,赵洪军  
【地点】：位于南安市石井镇林柄村的南安市定嘉石材厂厂棚  
【盗窃物品（价值）/现金】：该厂夹盘 22 个（价值人民币 658 元）  
【价值补充】：

#####

【时间】：2013 年 8 月 19 日 13 时许  
【作案人】：田光军,赵洪军  
【地点】：位于南安市石井镇郭前村的南安市石井三佳石材厂大切车间  
【盗窃物品（价值）/现金】：夹盘 17 个（价值人民币 595 元）  
【价值补充】：

\*\*\*\*\*

叶光平盗窃案 1C99FBBA-BA57-4E3C-B868-561D9EE7CAD4 false  
accusal: 0010502\_盗窃罪,  
clause: 第六十五条, 第六十七条, 第二百六十四条, 第六十四条,  
叶光平盗窃案  
上海市杨浦区人民法院  
刑事判决书  
(2014) 杨刑初字第 22 号  
公诉机关上海市杨浦区人民检察院。

#####

【时间】：2013 年 10 月 22 日 8 时许  
【作案人】：叶光平  
【地点】：本市杨浦区长海路某某某号长海医院健康体检中心  
【盗窃物品（价值）/现金】：HTC 牌 T328t 型手机一部

【价值补充】： 价值人民币 420 元

\*\*\*\*\*

符剑锋盗窃案 CBB92D4C-BF07-4F78-AEFC-0461E93FD544 true  
accusal: 0010502\_盗窃罪,  
clause: 第五十三条, 第六十七条, 第四十二条, 第二百六十四条, 第四十四条, 第五十二条,  
符剑锋盗窃案  
湖南省桃江县人民法院  
刑事判决书  
(2014)桃刑初字第 15 号  
公诉机关桃江县人民检察院。

#####

【时间】： 2013 年 9 月份的一天  
【作案人】： 符剑锋  
【地点】： 桃江县桃花江镇意中商店  
【盗窃物品（价值）/现金】： 香烟 8 包  
【价值补充】： 价值人民币 287 元

#####

【时间】： 2013 年 9 月份的一天  
【作案人】： 符剑锋  
【地点】： 桃江县桃花江镇凯旋汽车玻璃店  
【盗窃物品（价值）/现金】： 黄遇放在门前的黑色女式摩托车  
【价值补充】： 价值人民币 500 元

#####

【时间】： 2013 年 10 月 1 日  
【作案人】： 符剑锋  
【地点】： 桃江县桃花江镇洞庭美食城  
【盗窃物品（价值）/现金】： 周双喜放在门前的轩博马电动车一台  
【价值补充】： 价值人民币 1300 元

#####

【时间】： 2013 年 10 月 13 日  
【作案人】： 符剑锋  
【地点】： 桃江县桃花江镇步行街  
【盗窃物品（价值）/现金】： 现金 750 元  
【价值补充】：

\*\*\*\*\*

刘家安盗窃案 240321A8-D1C6-4236-9041-8BD3DAEFC444 true  
accusal: 0010502\_盗窃罪,

clause: 第五十三条, 第二百六十四条, 第五十二条,

刘家安盗窃案

柘城县人民法院

刑事判决书

(2013) 柘刑初字第 207 号

公诉机关柘城县人民检察院

#####

【时间】：	2013 年 7 月 4 日 8 时许
【作案人】：	刘家安
【地点】：	柘城县申桥乡孙庙村窑厂打工期间
【盗窃物品（价值）/现金】：	窑厂厂主张金花的一辆电动三轮车
【价值补充】：	价值 5951 元

\*\*\*\*\*

## 存在的问题

- 极个别文书存在编码问题，例如标点符号仍处于编码的状态，而不是标点符号本身；相关文书未处理
- 难以处理犯罪过程较为复杂的案件，例如诈骗形式的盗窃，文书中就会有长篇大论的对于诈骗过程的描述；相关文书未处理
- 难以处理没有显式的“盗窃”等相关字眼的文书，例如某案件中，说到“被告人将 XXX 车辆推走”，这种情况千变万化，不太好处理；相关文书未处理
- 难以处理写作形式不规范的文书；相关文书未处理
- 物品及其价值的对应存在问题。有一些估值是直接跟在物品后面的，所以就和物品连着输出了；但是有些没有连着，所以放在了[价值补充]中输出
- 物品的定语没有去除
- 文书本身对于时间的描述没有统一格式，且很多案件中，时间是模糊而不精确的，包含“左右”，“许”，“某天”等估计词，所以我也没有进行处理。

最终在 438214 篇文书中，成功提取了约 40 万篇文书。但是提取的信息也可能不准确，仍存在不达意、对不上号的情况。