**Report 1**

Summary：该程序用于<https://clinicaltrials.gov> 网站临床试验数据的爬取

1. 根据关键词检索 （term = cancer）

getUrl('https://clinicaltrials.gov/ct2/results',

{'term': 'cancer','currentpage': '1', 'pagesize': '5'})

2. 从所得html结果中，直接提取href，并进一步进行details的检索

eg.

href = ‘/ct2/show/NCT02053662?term=cancer&amp;rank=3’

变换成URL = ‘https://clinicaltrials.gov/ct2/show/record/NCT02053662?term=cancer&amp;rank=3’

(url = 'https://clinicaltrials.gov/ct2/show/record' + urlList[i][9:])

3. 用完整url获取details

getDetail(url)

4. 解析结果并保持：直接提取‘td’

说明：完整的数据有50多个变量，目前只选择了前4项保存到Excel中，后面可以根据需要进行修改