

上机作业： Shell脚本程序设计

北京邮电大学计算机学院 蒋砚军
jiangyanjun0718@bupt.edu.cn

第1题 生成TCP活动状况报告

生成TCP活动状况报告

netstat --statistics命令可以列出tcp等协议的统计信息。编写shell脚本程序，每隔1分钟生成1行信息：当前时间；这一分钟内TCP发送了多少报文；接收了多少报文；收发报文总数；行尾给出符号+或-或空格（+表示这分钟收发报文数比上分钟多10包以上，差别在10包或以内用空格，否则用符号-）。

运行示例如下：

| | | | | |
|------------------|------|------|------|---|
| 2020-04-17 00:02 | 345 | 314 | 659 | |
| 2020-04-17 00:03 | 1252 | 1100 | 2352 | + |
| 2020-04-17 00:04 | 714 | 570 | 1284 | - |
| 2020-04-17 00:05 | 151 | 139 | 290 | - |
| 2020-04-17 00:06 | 1550 | 1097 | 2647 | + |
| 2020-04-17 00:07 | 1385 | 959 | 2344 | - |
| 2020-04-17 00:08 | 5 | 1 | 6 | - |
| 2020-04-17 00:09 | 5 | 1 | 6 | |
| 2020-04-17 00:10 | 837 | 723 | 1560 | + |
| 2020-04-17 00:11 | 22 | 22 | 44 | - |

第2题 下载bing图库中图片 (选作)

下载bing图库中图片

访问<https://bing.ioliu.cn/?p=36>可以看到bing图库第36页的内容（见下一PPT中的图片），这个Web页有多个图片小样，将鼠标放到某个小样上，如右上角，可见中文说明信息“**野花草甸上的一只欧亚雕鸮，德国莱茵兰-普法尔茨**”和日期信息**2019-08-03**，点击一下，此图片就可以下载。

编写脚本程序**bing.sh**，将图库中照片全部下载下来存放到本地**bing**目录，上面URL中**p=36**可以换成**p=126**可访问126号页面，每页有12个图，每个图的日期，中文说明信息和下载地址及文件名html文件中可提取。要求下载后的文件命名为“**日期 说明.jpg**”例如：

2019-08-03 野花草甸上的一只欧亚雕鸮，德国莱茵兰-普法尔茨.jpg



要求

■ 不重复下载已下载的图片

检查图片是否已下载，如果已下载，则不再下载

支持批量任务在被中断后再次启动程序可以从中断处继续

■ 考虑并发

启动多个进程（`nproc=5`）同时下载，可以加快任务完成的速度。

■ 考虑下载文件出现故障的情况

◆ 如果一个图片有5MB，接收1.5MB后网络断开，则残存一个不完整的图片文件。

避免这种现象发生的一种方法是`wget`下载时使用一个临时文件名。

判断`wget`是否成功，若成功则将文件改名为正式名称；若失败，删除临时文件。

临时文件名的选取要考虑的并发问题，至少不可以程序中写死一个文件名导致两进程因使用相同的临时文件名而失败。

2020-03-18 藏红花.jpg
2020-03-19 柬埔寨吴哥窟的日出.jpg
2020-03-20 一只苍鹭栖息在加利福尼亚索尔顿海的木桩上.jpg
2020-03-21 LoughriggTarn的日出, 英国英格兰湖区安布尔塞德.jpg
2020-03-22 斋浦尔琥珀堡附近当地妇女正在爬阶梯井, 印度拉贾斯坦邦.jpg
2020-03-23 雷尼尔山透镜状云的延时视频, 华盛顿.jpg
2020-03-24 黄喉蜂虎, 西班牙埃斯特雷马杜拉.jpg
2020-03-25 拉德克利夫图书馆, 英国牛津.jpg
2020-03-26 鸟瞰蒙特雷附近的大苏尔海岸线, 加利福尼亚州.jpg
2020-03-27 在玉兰种植园里盛开的杜鹃花, 南卡罗来纳州查尔斯顿.jpg
2020-03-28 悬崖边的福门托尔角灯塔, 西班牙马略卡岛.jpg
2020-03-29 一只经受暴风雨的冠山雀, 苏格兰.jpg
2020-03-30 中央公园, 纽约.jpg
2020-03-31 连接巴林托伊附近两处悬崖的Carrick-a-Rede索桥, 北爱尔兰安特里姆.jpg
2020-04-01 马拉克勒国家公园中一只隐身的chia象, 南非林波波河.jpg
2020-04-02 从庞塞德莱昂河口向灯塔望去, 佛罗里达州.jpg
2020-04-03 维多利亚瀑布上的彩虹, 赞比亚.jpg
2020-04-04 【今日清明】哀悼为抗击新冠肺炎牺牲的烈士和逝世的同胞.jpg
2020-04-04 【今日清明节】.jpg
2020-04-05 基西米湖中的绿色树蛙和紫色睡莲, 佛罗里达州.jpg
2020-04-06 盛开的樱花和姬路城, 日本姬路.jpg
2020-04-07 四月的满月从圣迈克尔山上升起, 英国康沃尔.jpg
2020-04-08 圣托里尼岛鸟瞰图, 希腊.jpg
2020-04-09 巴芬岛附近的一群独角鲸, 加拿大努纳武特.jpg
2020-04-10 大熊雨林中一只柯莫德熊幼崽跟它的同胞挤在一起, 加拿大.jpg
2020-04-11 拉斯梅德拉斯的古罗马金矿遗址, 西班牙莱昂.jpg
2020-04-12 柏林亚历山大广场的复活节装饰.jpg
2020-04-13 WatChaloemPhraKiat PhrachomklaoRachanusorn的浮庙, 泰国南邦.jpg
2020-04-14 伊斯塔帕海岸的热带斑海豚, 墨西哥.jpg
2020-04-15 Bhagsu的热带瀑布, 印度喜马偕尔邦.jpg
2020-04-16 天生桥国家保护区中的sipapu桥, 犹他州.jpg
2020-04-17 阿尔冈金省立公园中一只正在树枝上休息的披肩鸡, 加拿大安大略省.jpg
2020-04-18 优胜美地国家公园中的春季瀑布.jpg
2020-04-19 内斯特角灯塔上空的银河, 苏格兰斯凯岛.jpg
2020-04-20 MicheldeverWood的蓝铃花, 英国汉普郡.jpg
2020-04-21 黄石国家公园的大棱镜泉, 怀俄明州.jpg
2020-04-22 怀波瓦森林中一棵名为TeMatuaNgahere的巨型贝壳杉树, 新西兰北地.jpg
2020-04-23 正在游水的雄性王绒鸭, 挪威特罗姆斯-芬马克郡.jpg
2020-04-24 由哈勃太空望远镜拍摄的大麦哲伦星云.jpg
2020-04-25 福克兰群岛上的南跳岩企鹅.jpg
2020-04-26 红宝石海滩的日落, 华盛顿州奥林匹克国家公园.jpg
2020-04-27 一只南美獾幼崽小跑着穿过草地.jpg
2020-04-28 索尔兹伯里大教堂与放牧的羊群, 英格兰.jpg
2020-04-29 普尔曼附近的帕卢斯一辆拖拉机在耕作时扬起尘土, 华盛顿州.jpg

实验环境存在的问题

网站**bing.ioliu.cn**有访问限制，当该站点观察到一段时间内来自同一个IP地址的下载请求较多，站点将拒绝服务。过一段时间后恢复。

```
$ wget https://bing.ioliu.cn -O page.html
--2022-03-17 22:56:37-- https://bing.ioliu.cn/
Resolving bing.ioliu.cn (bing.ioliu.cn)... 120.78.199.148
Connecting to bing.ioliu.cn (bing.ioliu.cn)|120.78.199.148|:443... connected.
HTTP request sent, awaiting response... 403 Forbidden
2022-03-17 22:56:37 ERROR 403: Forbidden.
```

多位同学聚集在校园通过NAT访问这个网站，或者通过云端同一台Linux访问，都会遇到这样的问题。家庭用户可以控制路由器重新联网得到新IP地址以避开这个问题。问题无法解决的，可选做第3题，第3题的网站访问没有类似的限制。

程序代码量100行以内。

第3题 下载畅销榜上的小说

(选作)

人气榜单

热门作品排行

月票榜

畅销榜 ▶

阅读指数榜

粉丝榜

推荐榜

收藏榜

更新榜

VIP收藏榜

打赏榜

新书排行

签约作者新书榜

公众作者新书榜

新人签约新书榜

新人作者新书榜

社区排行

打赏粉丝榜

其他排行

女生精选榜

女生月票榜

畅销榜

本日作品销量排行

全部分类 · 玄幻 · 奇幻 · 武侠 · 仙侠 · 都市 · 现实 · 军事 · 历史 · 游戏 · 体育 · 科幻 · 诸天无限 · 悬疑 · 轻小说



夜的命名术

👤 会说话的肘子 | 都市 · 异术超能 | 连载

蓝与紫的霓虹中，浓密的钢铁苍穹下，数据洪流的前端，是科技革命之后的世界，也是现实与虚幻的分界。钢铁与身体，过去与未来。这里，表世界与里世界并存，面前的一

最新更新 677、众生相 · 2022-03-17 13:23

书籍详情

加入书架



星门

👤 老鹰吃小鸡 | 玄幻 · 高武世界 | 连载

传说，在那古老的星空深处，伫立着一道血与火侵染的红色之门。传奇与神话，黑暗与光明，无尽传说皆在这古老的门户中流淌。俯瞰星门，热血照耀天地，黑暗终将离去！

最新更新 再次延迟一下 · 2022-03-17 15:04

书籍详情

加入书架



不科学御兽

👤 轻泉流响 | 玄幻 · 异世大陆 | 连载

这是一个以御兽为主流的异世界。动物，植物，元素，死灵，机械……九尾炎狐，生命古树，雷云精灵，骷髅君主，机械舰娘乃至孕育出灵的古董刀剑皆可作为宠兽，而契

最新更新 第374章：蓝发… · 2022-03-16 22:11

书籍详情

加入书架



黎明之剑

👤 远瞳 | 科幻 · 时空穿梭 | 连载

高文穿越了，但穿越的时候稍微出了点问题。在某个异界大陆上空飘了几十万年之后，他觉得自己可能需要一具身体才算是成为一个完整的穿越者，但他没想到自己好不容

最新更新 完本感言 · 2022-03-16 11:03

书籍详情

加入书架

下载起点中文网畅销榜小说

访问<https://www.qidian.com/rank/hotsales>可以进入起点中文网畅销榜。将这个畅销榜上的所有小说下载下来，每部小说目录页面中有VIP标记需要订阅才能访问的章节略掉。将每部小说整理成一个文本文件，文本文件的命名为“**排行榜名次-小说的名字.txt**”。要求文本文件粘贴到Word中后可以很方便地生成章节目录，并方便正文排版。

程序代码量100行之内。



黎明之剑



模板



黎明之剑



黎明之剑

提示：整理文本文件时可使用ed在脚本程序中编辑文件，用于删除文件的一段

| | | |
|-------------------|--------------------|-----------------------|
| 0001-夜的命名术.txt | 0008-轮回乐园.txt | 0015-深空彼岸.txt |
| 0002-星门.txt | 0009-我用闲书成圣人.txt | 0016-术师手册.txt |
| 0003-明克街13号.txt | 0010-大奉打更人.txt | 0017-诸界第一因.txt |
| 0004-不科学御兽.txt | 0011-基因大时代.txt | 0018-我就是不按套路出牌.txt |
| 0005-黎明之剑.txt | 0012-我已不做大佬好多年.txt | 0019-女主从书里跑出来了怎么办.txt |
| 0006-家父汉高祖.txt | 0013-这个明星很想退休.txt | 0020-LOL：这个男人太强了!.txt |
| 0007-这游戏也太真实了.txt | 0014-诡秘之主.txt | |

要求

■ 不重复下载已下载的信息，支持任务被中断后可以从中断处继续

- ◆ 如果某部小说已经下载完毕，程序再次运行时不再下载
- ◆ 如果某部小说有90个页面构成，已经下载了47个，程序再次运行时从48开始

提示：整理文件时可以考虑用使用`ed`命令在脚本程序中编辑文件。

■ 考虑并发

- ◆ 启动多个线程（`nproc=5`）同时下载，可以加快任务完成的速度

■ 考虑下载文件出现故障的情况

- ◆ 如果一个html页面接收过程中网络断开，则残存一个不完整的html文件

避免这种现象发生的一种方法是`wget`下载时使用一个临时文件名。

判断`wget`是否成功，若成功则将文件改名为正式名称；若失败，删除临时文件。

临时文件名的选取要考虑的并发问题，至少不可以程序中写死一个文件名导致两进程因使用相同的临时文件名而失败。