

به نام خدا



تمرین شماره ۰ - تمرین لینوکس

تحلیل داده‌های حجیم زیستی (۲-۱۳۹۹) - دکتر مطهری

**\*\* برای تمرین‌های درس لطفا گزارش کاملی بنویسید.**

**\*\* برای گزارش تمرین شماره ۰ حتما از مراحل انجام کار خود عکس بگذارید.**

**\*\* برای کدها و فایل‌های هر تمرین پوشه‌ی جدا در نظر بگیرید. مثلا برای تمرین شماره صفر پوشه تمرین شما باید شامل پوشه سوال اول، پوشه سوال دوم و گزارش تمرین باشد.**

**\*\* فایل نهایی تمرین را zip کرده و به این صورت ارسال نمایید: `HW[number]_[student_number].zip`**

**\*\* در صورت همکاری در تمرین نام دوستان خود را در گزارش تمرین بنویسید ☺**

**\*\* اگر برای این تمرین دچار مشکل در نصب سیستم شده‌اید می‌توانید از سرور دانشکده استفاده کنید.**

**\*\* سوالات خود را می‌توانید به این آدرس ارسال نمایید : `sepideh_m2006@yahoo.co.uk`**

**\*\* مهلت ارسال تمرین : ۴ اسفند ۹۹**

## سوال ۱:

در ابتدا فایل [Mesculenta\\_147\\_gene.gff3.gz](http://Mesculenta_147_gene.gff3.gz) را در دستگاه خود دانلود کنید و عملیات زیر را بر روی آن انجام داده و عکس خروجی‌های خود را در گزارش بگذارید:

۱. با دستور `gunzip` فایل را باز کرده و نام آن را به `Cassavagenes.txt` تغییر دهید.

۲. تعداد ژن‌های متفاوت در فایل را با دستور `grep` پیدا کنید.

۳. سطرهایی که دارای کلمه `gene` است را در یک فایل به نام `Genes.txt` ذخیره کنید.

۴. تعداد ژن‌های موجود در فایل `Genes.txt` را با دستور `wc` بیابید.

۵. تا جایی که می‌توانید در این دستورات از `pipe` استفاده کنید.

## سوال ۲:

ابتدا فایل [hw0\\_2.tar.gz](http://hw0_2.tar.gz) را با دستور `wget` دانلود نمایید. پس از `decompress` کردن فایل با دستور `tar`، در پوشه `hw0_2` تعدادی پوشه با نام‌های مختلف می‌بینید. درون هر پوشه اطلاعات یک ژن قرار گرفته است. اطلاعات هر ژن شامل نام، محل قرار گرفتن آن بر روی `srand` و طول رشته آن است. در این سوال قصد داریم با نوشتن `bash script` ای ابتدا از هر فایل اطلاعات آن را خوانده و برای هر ژن رشته رندمی از `ATCG` ایجاد کرده و نتیجه آن را به همراه اطلاعات ژن در یک فایل دیگر که نام آن را از

ورودی دریافت کرده ذخیره کنیم. در فایل خروجی در هر سطر اسم ژن، محل قرارگیری، رشته رندم تولید شده به اندازه طول آن، قرار گرفته و ترتیب قرارگیری ژن ها باید براساس طولشان مرتب شده باشد. (لطفا در گزارش عکس گام به گام فراموش نشود 😊)

```
sepideh@sepideh:~/Desktop/hw0_2$ ls
gene1 gene10 gene2 gene3 gene4 gene5 gene6 gene7 gene8 gene9 test.sh
sepideh@sepideh:~/Desktop/hw0_2$ ./test.sh gene.out
sepideh@sepideh:~/Desktop/hw0_2$ cat gene.out
tvsa - GATGC
xgo + TACTA
ldt4 - ACCGAT
pixrs1fg - ATTCCTT
xbk + TAGTGTCGA
a1 + AATCACTAAG
xti - TATACTAAAA
lpm + TAGCAAATGGT
dfg - ACGCTTCTTAAG
spvrtyv + GTTCACGGACAATTC
sepideh@sepideh:~/Desktop/hw0_2$
```

**\*\* راهنمایی که ممکنه به کارتون بیاد:**

برای چاپ کردن ستون های دلخواه خود می توانید از دستور awk استفاده کنید.  
از دستور printf نیز برای چاپ خط به خط خروجی در فایل می توانید استفاده کنید.