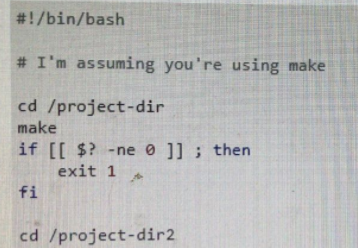
**מתמ – HW1 : תכנון תרגיל:**

לתרגיל יש 3 סקריפטים – כלומר 3 קבצים.

**הערות לשתינו:**

* יציאה מהסקריפט : לכתוב את הפקודה exit.



**סעיף א':**

**שם הסקריפט:** rawFileFilter

מהלך עבודה:

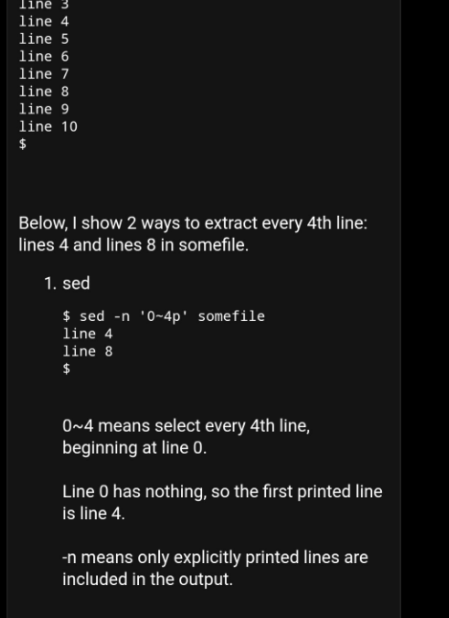
ההתייחסות לפרמטרים, תיעשה כ:

Scriptname param1 param2 param3

ובתוך התוכנית מתייחסים אליהם כ $1 $2 $3 .

מומלץ בתחילת הסקריפט לעשות נגיד file=$1.

* סינון הקובץ לפי שורות ה DNA.



* להוציא את כל התווים הלא חוקיים. להשאיר בכל שורה רק את התווים החוקיים.

להשתמש בדוגמה מס 20

**20. To remove a character irrespective of the case**:

$ sed 's/[uU]//g' file

* אחרי שמסננים צריך להוציא את כל התווים פרט ל num התווים הראשונים. משתמשים בדוגמה מס 10 בפקודה sed:

**10. To remove everything except the 1st n characters in every line**:

$ sed -r 's/(.{3}).\*/\1/' file

* להעביר את השורות המתאימות לקובץ הפלט outfile. כל שורה אחת אחרי שנייה.
* צריך לשים את הקובץ פלט לתיקייה בשם num. ע"י פקודה mov או cp .

**מקרי קצה:**

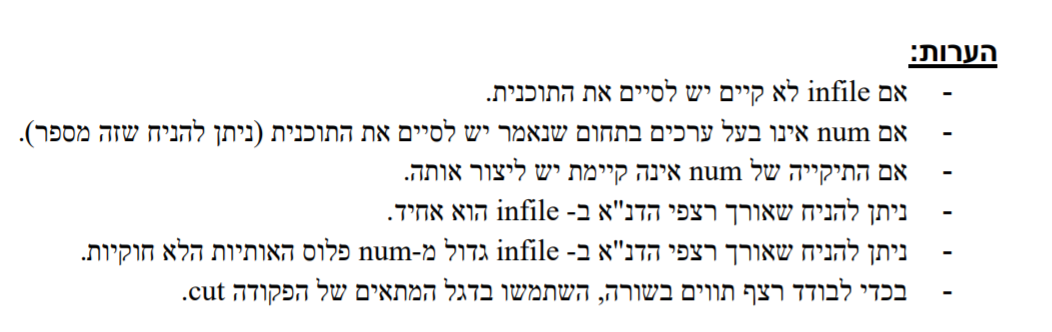
1. **שאלה מהמודל:**

יצירת outfile בסעיף א':

האם הכוונה היא ליצור קובץ outfile עם מספר סידורי כלשהו לכל הרצה של הסקריפט על קבצי קלט? כלומר אם אני מפעיל את הסקריפט על file1.fastq, ואז על file2.fastq ואז שוב על file1, יווצרו לי 3 קבצי out בתיקיה num? (בהנחה שבשלושת הפעמים קראתי לסקריפט עם אותו num)

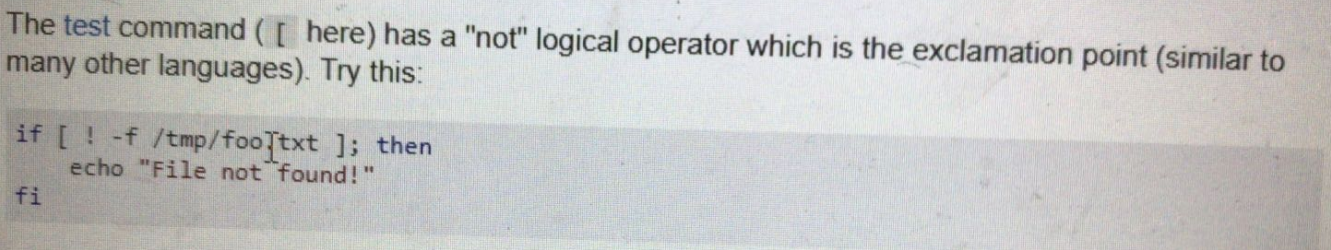
התשובה:

לא. במידה וקיים כבר הקובץ outfile בתיקייה הספציפית אז יהיה overwrite לערכו.

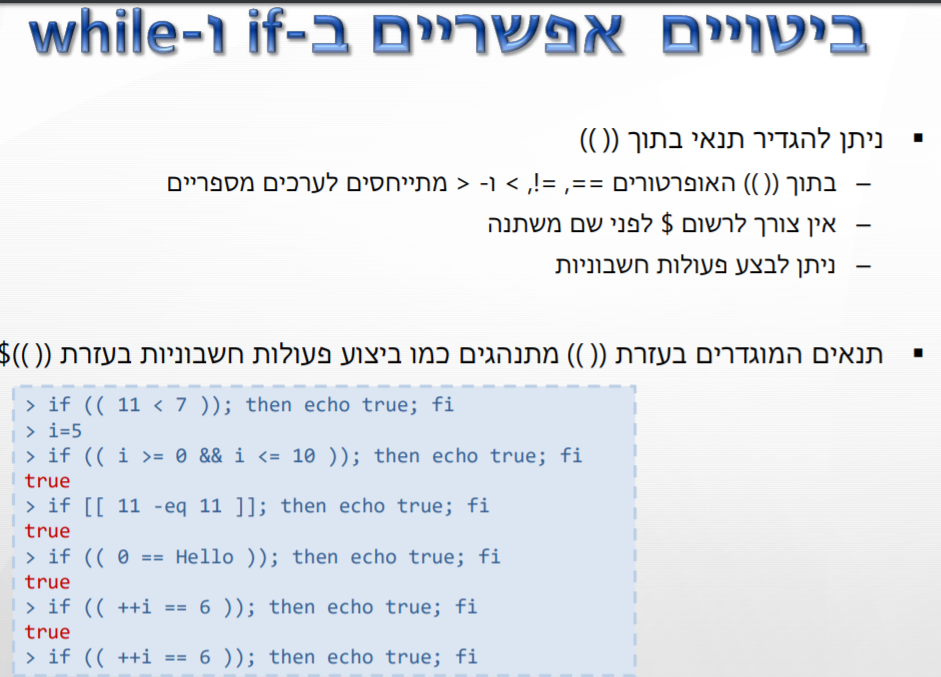
1. לשים לב שיכולות להיות מלא שורות ריקות בין רצפים.

פתרון למקרי קצה:

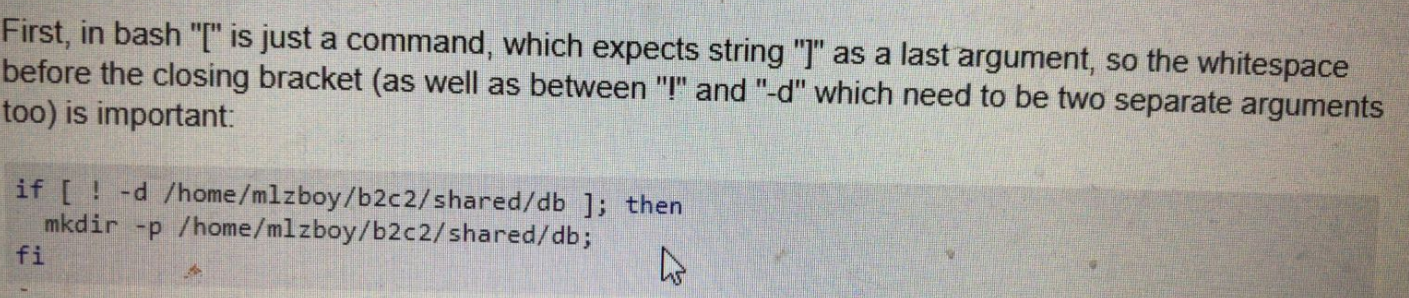












**סעיף ב':**

**שם הסקריפט:** count

מהלך עבודה:

ההתייחסות לפרמטרים, תיעשה כ:

Scriptname param1 param2 param3

ובתוך התוכנית מתייחסים אליהם כ $1 $2 $3 .

מומלץ בתחילת הסקריפט לעשות נגיד file=$1.

* לאגד את כל התכנים שבכל הקבצים למקום אחד.

לחשוב לאן לאגד אותם! כי אסור קבצים זמניים.

להשתמש בפקודה:

cat \* |

(צריך לשים לב להיות בתיקייה הרלוונטית)

* למיין לפי מספר הופעתם בסדר יורד, כאשר במקרה תיקו, ההופעה תיהיה לפי סדר לקסיקוגרפי יורד.

להשתמש בפקודות sort עם התייחסות לסדר יורד ( דגל -r) , ולהמשיך בפייפלין uniq (עם דגל -c).

* למחוק את השורה האחרונה – שורת הרווחים : בעזרת פקודת head -n-1



* דרוש מיון נוסף לפי מספר הופעתם בסדר יורד.

משתמשים ב sort (עם דגלs -n r) בשביל לשמור על מיון יציב !!

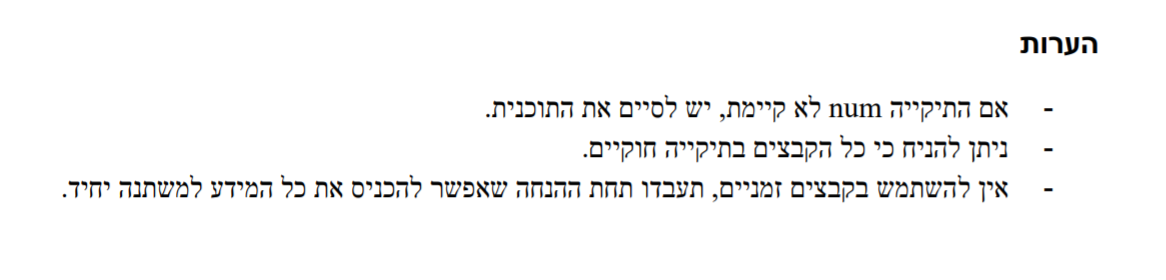
* הדפסת התוצאה:

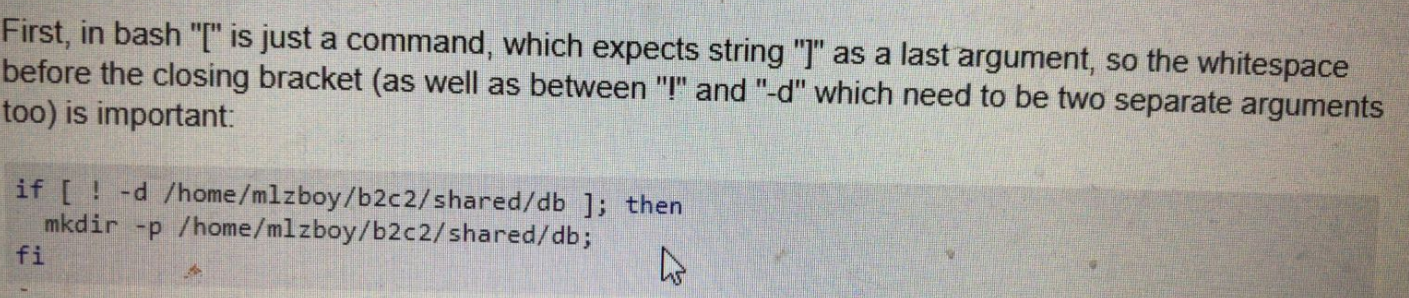
בעזרת echo. הטאב ייעשה באמצעות גרשיים.

מקרי קצה:

1. לשים לב למקרה של תיקו.

countיאגד את כל הקבצים שבתיקייה num ואז ימיין את הרצפים ע"פ מספר הופעתם בסדר יורד כאשר במקרה תיקו ההופעה לפי סדר לקסיקוגרפי יורד ויוציא אותם כפלט סטנדרטי בפורמט הבא.





**סעיף ג':**

**שם הסקריפט:** seperateSeq

**מהלך עבודה:**

קלט: infile: שם של קובץ. זהו הפלט של סעיף א'.

**מהלך עבודה:**

* ההתייחסות לפרמטרים, תיעשה כ:

Scriptname param1 param2 param3

ובתוך התוכנית מתייחסים אליהם כ $1 $2 $3 .

מומלץ בתחילת הסקריפט לעשות נגיד file=$1.

* להגדיר משתנה line – שיהיה שווה למספר השורות שבקובץ. ע"י הפקודה:
* for (( i=0; i<${#line}; i++ )); do  
          if [[ "${line:$i:1}" == "T" ]]; then

מקרי קצה:

* שם הקובץ יכול להכיל רווחים!!! צריך להתייחס לזה.
* אם infile לא קיים יש לסיים את התוכנית:

