



הנדסת תוכנה 094129

תרגיל בית 1

עבודה עם String, Vector ומערכים.

תאריך אחרון להגשה: 9/5/2019 בשעה 23:55

מתרגל אחראי: רום גוטמן

נושא התרגיל: מערכות מידע גנטיות

תיאור המודל הביולוגי:

במסגרת עבודתכם כמפתחים מתמחים בחברת ביוטק נודעת, כחלק מקבוצת חוקרי ריצוף גנום מובחרים, נבחרתם לפתח מערכת מידע גנטית. ה-DNA (Deoxyribonucleic Acid) הינה מולקולת ענק של חומצות גרעין הבנויה כשרשרת נוקליאוטידים. כל נוקלאוטיד מיוצג כאות מהא"ב $\{G, T, A, C\}$. כידוע ה-DNA מקודד לחלבונים הבונים את הגוף ומסייעים לתחזוקתו. החלבונים בנויים כשרשרת של חומצות אמיניות, אותן ניתן לייצג כמחרוזת המייצגת חומצה אמינית כלשהי. ידוע כי רצף של שלושה נוקלאוטידים מקודד לחומצה אמינית כלשהי (כאשר הרצף נקרא משמאל לימין), שימו לב כי ניתן לקודד לחומצה אמינית מסוימת יותר מדרך אחת.

סדרה של שלשות של נוקלאוטידים אשר מקודדות לחלבון מסוים נקראת גן. למשל אם רצף הנוקלאוטידים GCC מקודד לחומצה האמינית a ו-AAA מקודד לחומצה האמינית h ו-CGT מקודד לחומצה האמינית r אז הגן GCC-AAA-CGT מקודד לחלבון a-h-r. עקב תהליכים אבולוציוניים מורכבים, הגנים לא נמצאים באופן רציף ב-DNA ויתכן כי הם יהיו מפוזרים לאורכו.

עליכם הוטלה המשימה לחלץ את הגנים מתוך DNA נתון, בעזרת מיפוי וסדר החלבון הנדרש.

מבנה הקלט וכללי הפיענוח:סוגי הקלט:

1. **מחרוזת (string) ה-DNA** אשר מורכב מאותיות הא"ב {G,T,A,C} בלבד.
 2. **מחרוזת המיפוי**, מהצורה הכללית $i_{1,1}, i_{1,2}, i_{1,3} Acid_1 \dots i_{n,1}, i_{n,2}, i_{n,3} Acid_n$ כאשר i מייצג אינדקס (מספר אי-שלילי) בשרשרת ה-DNA, אשר מרכיב את חומצה $Acid_j$ ו- j מייצג שם של החומצה j (מחרוזת). שימו לב כי יש רווח בין האינדקסים לבין השמות וכי האינדקסים מופרדים בפסיקים.
 3. **מחרוזת ההדפסה** - הגן אותו נרצה לקבל: קלט מהצורה $Acid_1 Acid_2 Acid_3 \dots Acid_k$ (שימו לב כי המחרוזות מופרדות ברווח)
 4. **מחרוזת הסיום** - המילה "end"
- בתוכנה שלכם, עליכם לצפות לקבל כקלט את **מחרוזת ה-DNA** ולאחריה את **מחרוזת המיפוי**. לאחר קבלת שני הקלטים, עליכם לצפות לקבל מספר לא ידוע של **מחרוזות הדפסה**, עד אשר תקבלו את **מחרוזת הסיום** אשר תסיים את פעולת התוכנה. (ראו דוגמא בהמשך)
- הפיענוח כולל 3 שלבים:

1. החלפת האינדקסים ב**מחרוזת המיפוי** לאותיות הנמצאות באינדקס הנתון ב**מחרוזת ה-DNA** ובכך לקבל את **מחרוזת הגן** המתאימה.
2. נדרש להחליף כל שלשה רציפה ב**מחרוזת הגן** בשם **החומצה** המייצגת את החומצה האמינית המתאימה.
3. הדפסת החלבון במקודד ע"י **מחרוזת ההדפסה** הנתונה.

הפלט של המערכת:

דוגמא לאינטרקציה עם המערכת. **קלט** באדום, **פלט** בירוק:

```

TAGCGCGTACCTAGCATAc
1,3,0 a 11,2,2 b 5,0,1 c 4,13,6 Val
c a b
CTAACTTGG
b c c a
TGGCTACTAACT
Val a c
GGGACTCTA
end

```

הנחות שמותר לכם להניח לגבי הקלט:

- ניתן להניח תקינות קלט.
- ניתן להניח שהטקסט שמתקבל הוא ע"פ הא"ב שהוגדר בשאלה.
- בין אותיות הנוקלאוטידים במחרוזת ה-DNA אין רווח.
- האינדקסים במחרוזת המיפוי מופרדות בפסיק ";", ". האינדקסים ושם החומצה במחרוזת זו מופרדים ברווח.
- המחרוזות במחרוזת ההדפסה מופרדות זו מזו על ידי רווח בודד.

דרישות המערכת:

- במימוש עליכם להשתמש בכל אחת מהטיפוסים הבאים:
`string`, `vector`, `iterator`, `const_iterator`
- עליכם לשמור על כללי תכנות נאות כפי שפורסם במסמך במודל, ובפרט להקפיד על:
 1. אתחול משתנים ולתת להם שמות משמעותיים.
 2. לא לכתוב ליטרלים בקוד (אלא להגדיר קבועים מתאימים).
 3. פרט לקיצורים וראשי התיבות המופיעים במסמך, ראשי תיבות **לא** נחשב לשם בעל משמעות
- עליכם להשתמש בדוגמאות מההרצאות והתרגולים. כמו כן אתם רשאים להשתמש בפונק' נוספות של טיפוס הספריות הנ"ל שלא נלמדו.
- עליכם לתעד את הקוד שלכם באנגלית בלבד (אין צורך לדאוג לרמת האנגלית). יש לתעד לפי סגנון התיעוד שמופיע במודל.
- אין בתרגיל זה דרישה של יעילות זמן.
- עבודה עם (בפרט על) מיכלים סדרתיים תעשה אך ורק בעזרת איטרטור המתאים ביותר.
- הקדישו תשומת לב ומחשבה לאופן בו הפונקציות שלכם מקבלות פרמטרים.
- נדרש מכם לכתוב את הקוד בקובץ `main.cpp` ואותו הגישו, זאת מכיוון שטרם למדתם איך מחלקים את הקוד בצורה נכונה (אנו ממליצים לכם להתנסות בחלוקה ושימוש נכון במודלים, כפי שנגענו בקצרה בתרגולים).
- כמו כן – אנו נותנים לכם חתימות של פונקציות, אך במידה ואתם רוצים – אתם רשאים לבצע חלוקה כזו לבד ולא להשתמש במימוש שלנו.
- **יש להקפיד כי פלט הקבצים זהה לחלוטין לפלט הניתן לכם.**

הוראות הגשה:

1. התרגיל להגשה בזוגות **בלבד**.
2. הקוד חייב להיכתב על פי מסכמות כתיבת הקוד (Coding Conventions) בקורס. קוד של עומד במסכמות לא יזכה במלוא הניקוד.
3. חובה לקמפל את התרגיל בעזרת סביבת TDM GCC שעליכם להתקין על פי המדריך- תוכנית אשר תצליח להתקמפל בסביבה אחרת ולא בסביבה זו **תחשב כקוד של עובר קומפילציה**.
4. ההגשה חייבת להכיל קובץ **ZIP יחיד בלבד** (ולא קובץ RAR וכדומה) המכיל:
 - תיקייה בשם code ובה כל קבצי קוד המקור (h/cpp), ללא קבצי ההרצה (.exe).
5. שם הקובץ חייב להיות hw1_xxxxxxxx_yyyyyyyyyy.zip כאשר xxxxxxxx ו- yyyyyyyyyy הם מספרי תעודות הזהות של המגישים כולל ספרת ביקורת.
6. ההגשה היא אלקטרונית **בלבד**, דרך אתר ה-moodle של הקורס. תרגילים שיוגשו בכל דרך אחרת **לא ייבדקו**.
7. אין להגיש את אותו הקובץ פעמיים. התרגיל יוגש על ידי **אחד** מבני הזוג.
8. תרגיל בית של יוגש על פי הוראות ההגשה – לא ייבדק. **(כלומר ציון 0)**
9. יש להקפיד על יושרת הכנת התרגיל וההגשה.
10. קוד אשר לא יעבור קומפילציה יזכה לציון 0.

בהצלחה!

.....