

# תרגיל בית מספר 1

## נושא: סימולציית נגיפי קורונה בעזרת STL ותמיכת Reference Counting

דדליין: יום א', 24/04/2022, 23:59

הגשה ביחידים

בהצלחה רבה!

### תיאור התרגיל

בתרגיל זה תממשו עולם מונחה עצמים של אוכלוסיית נגיפי קורונה, כאשר נגיף מוגדר באמצעות מחרוזת תווים בעלת אלפבית ייחודי, ותכונותיו יתוארו בהרחבה בהמשך. מטרת התרגיל היא מימוש של סימולציה תלוית-זמן תוך שימוש נרחב בספריית STL וכן תמיכה ב-*reference counting*.

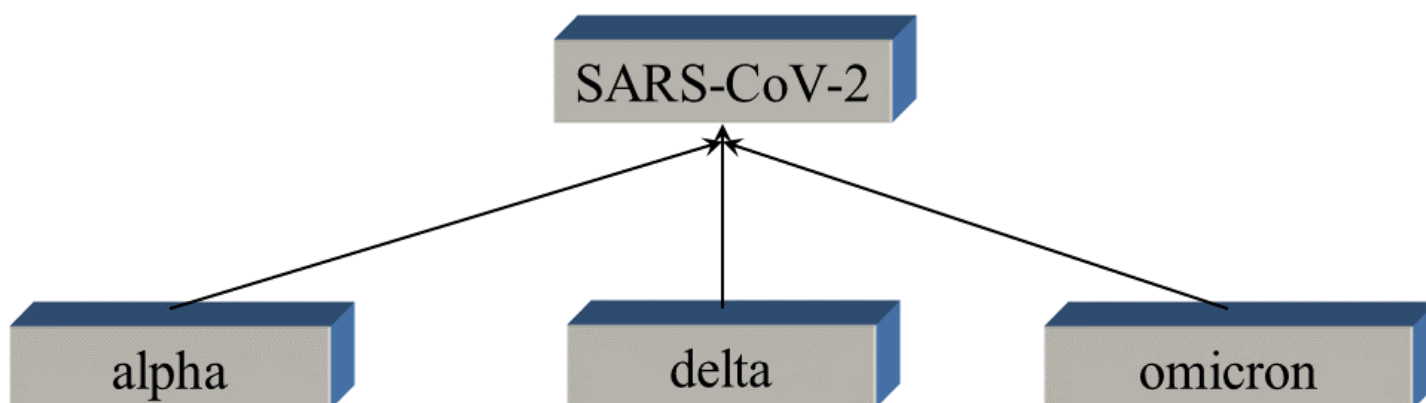
### אובייקט נגיף: אפיון

נגיף הינו אובייקט בעולם הסימולציה אשר מאופיין באמצעות מחרוזת תווים באורך קבוע המבוססת על האלפבית הבא: {A,T,C,G}. ההנחה היא כי בעת בדיקת מצב העולם (כלומר בכל פעימת זמן בדידה, או דור) קיימת אוכלוסייה בגודל קבוע של נגיפים, כל נגיף מכיל מחרוזת חוקית, וכי הם מבצעים עדכוני מחרוזות בין נקודת זמן אחת לאחרת באופן שיפורט.

עם הגדרת העולם, יוגדר **הדור ההתחלתי** של הנגיפים; מחרוזות התווים של דור זה נחשבות ל"אבות הקדמונים" של המחרוזות במהלך הסימולציה. בפרט, על כל נגיף לזכור במהלך הסימולציה מיהו "האב הקדמון" של מחרוזת התווים שלו – ועל זיכרון זה להיות ממומש באמצעות מנגנון *reference counting*.

עם הגדרת העולם, תוגדר גם **מחרוזת מטרה**, אליה "שואפת להגיע" אוכלוסיית הנגיפים. בכל נקודת זמן ניתן לחשב את מספר הפגיעות של כל נגיף ביחס למחרוזת המטרה (כלומר, מספר הקואורדינטות במחרוזת הנגיף בהן התו זהה לתו המטרה באותו מיקום), ומכאן ניתן להגדיר את עוצמת ההתאמה של נגיף נתון ביחס לנגיף המטרה. הנגיף הדומה ביותר למחרוזת המטרה, כלומר בעל מידת ההתאמה החזקה ביותר, ייחשב לנגיף החזק באותה נקודת זמן (*fittest*); באופן דומה, הנגיף בעל השוני הרב ביותר ביחס למחרוזת המטרה ייחשב לנגיף החלש ביותר באוכלוסייה.

הנגיפים בהם הסימולציה תתחשב נקראים אלפא ('a'), דלתא ('d'), ואומיקרון ('o'). מבחינת תיאור אובייקטים, שלושתם שייכים לממלכת נגיפי הקורונה (יורשים מאובייקט "נגיף קורונה", SARS-CoV-2). עץ הירושה המתאר ממלכה זו נתון בסכימה הבאה:



# נושאים מתקדמים בתכנות מונחה-עצמים, אביב 2022

## עוצמת ההתאמה של נגיף נתון ביחס למחרוזת המטרה

יהי מימד מחרוזות התווים של הנגיפים  $\ell$ , ותהי  $\vec{a}$  מחרוזת חוקית של נגיף נתון המקיימת

$$\vec{a} \in \{A, T, C, G\}^\ell$$

תהי  $\vec{\tau}$  מחרוזת המטרה המקיימת גם-כן  $\vec{\tau} \in \{A, T, C, G\}^\ell$ . נגדיר את מספר הפגיעות של המחרוזת  $\vec{a}$  במחרוזת המטרה  $\vec{\tau}$  כמספר הקואורדינטות בהן מתקיים שוויון תווים:

$$h(\vec{a}, \vec{\tau}) = \left| \{k \mid a_k = \tau_k, k = 1 \dots \ell\} \right|$$

בהתאם, **עוצמת ההתאמה** של הנגיף הנושא את המחרוזת  $\vec{a}$  ביחס למחרוזת המטרה  $\vec{\tau}$  הינה סקלר ב-[0,1] המוגדר כמספר הפגיעות המנורמל באורך המחרוזת:

$$f_{\vec{\tau}}(\vec{a}) = \frac{h}{\ell} \in [0,1]$$

אחד מתנאי העצירה של הסימולציה הוא הימצאותה של מחרוזת בעלת עוצמת התאמה 1.0 אצל אחד מהנגיפים באוכלוסייה.

## עדכון נגיפים במעבר דור

עדכון הנגיפים מתרחש בכל פעימת זמן בדידה (דור), והוא **הסתברותי** בטבעו. קיימים שני סוגי עדכון – עדכון פרטני, ועדכון קבוצתי – כפי שיוסבר בפירוט בסעיף זה.

**עדכון פרטני:** בהינתן הסתברות שינוי  $p_m$ , כל קואורדינטה במחרוזת תוחלף בהסתברות זו, כאשר אלו הם השינויים המותרים:

$$A \rightarrow T \quad T \rightarrow A \quad C \rightarrow G \quad G \rightarrow C$$

הסתברות השינוי  $p_m$ , הינה קבועה עבור כל אחת מן המשפחות ושינויי המחרוזות המותרים נקבעים ע"פ משפחת הנגיף:

- למשפחת אלפא תהיה הסתברות בעלת ערך  $p_m^s = 0.37$  והשינויים המותרים הינם

$$A \rightarrow T \quad T \rightarrow A$$

- למשפחת דלתא תהיה הסתברות בעלת ערך  $p_m^m = 1/\ell$  והשינויים המותרים הינם

$$A \rightarrow T \quad C \rightarrow G \quad G \rightarrow C$$

- למשפחת אומיקרון תהיה הסתברות בעלת ערך  $p_m^c = 2/\ell$  והשינויים המותרים הינם

$$T \rightarrow A \quad C \rightarrow G \quad G \rightarrow C$$

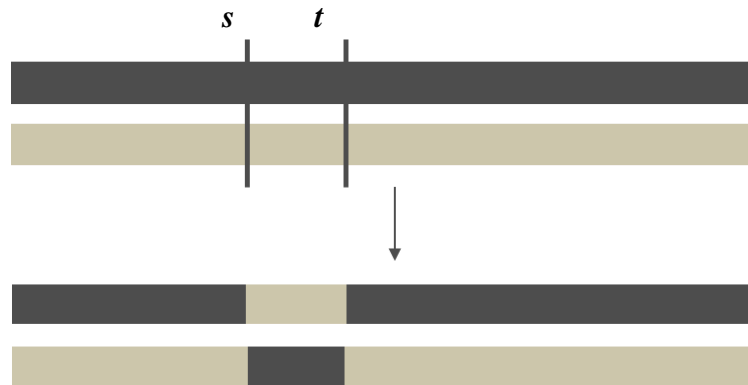
**עדכון קבוצתי:** בכל דור ייבחר באופן אקראי צמד יחיד של נגיפים מבין **הדור ההתחלתי**, אשר יבצעו ביניהם החלפה של מקטעי מחרוזת. לאחר שנבחר זוג נגיפים, יוגרלו שני אינדקסים אקראיים בהם יתרחשו חיתוכי מחרוזות לשם החלפת מקטעי תווים. באופן מפורש, נסמן ב- $i$  את זהות הנגיף הראשון, וב- $j$  את זהות הנגיף השני, ונתייחס אל  $\vec{a}^{(i)}, \vec{a}^{(j)}$  כאל מחרוזות התווים של נגיפים אלו, בהתאמה. יהיו  $\{s, t\}$  האינדקסים לחיתוך המקטעים,  $s < t$ ,  $s \in \{2, \dots, \ell-2\}$ ,  $t \in \{3, \dots, \ell-1\}$ . לאחר פעולת ההחלפה, כך תיראנה המחרוזות:

$$\vec{a}_{XO}^{(i)} = (a_1^{(i)}, \dots, a_s^{(i)}, a_{s+1}^{(j)}, \dots, a_t^{(j)}, a_{t+1}^{(i)}, \dots, a_\ell^{(i)})$$

$$\vec{a}_{XO}^{(j)} = (a_1^{(j)}, \dots, a_s^{(j)}, a_{s+1}^{(i)}, \dots, a_t^{(i)}, a_{t+1}^{(j)}, \dots, a_\ell^{(j)})$$

# נושאים מתקדמים בתכנות מונחה-עצמים, אביב 2022

התרשים הבא נועד להמחיש את פעולת יצירת המחרוזות החדשות:



במקרה זה נאמר שהמחרוזת  $\vec{a}_{xo}^{(i)}$  מקורה באב הקדמון  $i$  ונגדיר את הנגיף שלה להיות ממשפחת הנגיף  $i$ , והמחרוזת  $\vec{a}_{xo}^{(j)}$  מקורה באב הקדמון  $j$  וכי הנגיף שלה הוא ממשפחת הנגיף  $j$ .

בפעולת העדכון הקבוצתי, שתי המחרוזות החדשות  $\vec{a}_{xo}^{(i)}, \vec{a}_{xo}^{(j)}$  תחלפנה את שתי המחרוזות החלשות ביותר באוכלוסייה.

הרעיון הכללי של התכנית אותה תממשו הוא לקלוט מן המשתמש אוכלוסיית נגיפים בתור דור התחלתי, ולסמלץ את תוכן המחרוזות, במעברי זמן בדידים, עד להגעה של הנגיף הראשון למחרוזת המטרה (עוצמת התאמה 1.0), אז תסתיים הסימולציה. בכל מעבר זמן (דור), תעודכנה מחרוזות הנגיפים לפי מתכונת העדכון ההסתברותית שלהן, ובכל נגיף יהיה זיכרון של מחרוזת "האב הקדמון" ממנה הוא נגזר. בסופה של התכנית, הפלט יכלול את מחרוזות הנגיפים של הדור האחרון, את מניית המופעים של כל "מחרוזת אב קדמון", וכן את מחרוזת הנגיף החזק ביותר בהיסטוריית הריצה.

# נושאים מתקדמים בתכנות מונחה-עצמים, אביב 2022

## פורמט קלט

התכנית תקבל כקלט שני קבצים – הראשון יתאר את עולם הנגיפים (פרמטרים שונים, מחרוזת המטרה וכיו"ב), והשני יכלול את הדור ההתחלתי. לאחר קומפילציה של התכנית לכדי קובץ הרצה בשם simCOVID, כך ניתן יהיה להריץ אותה במקרה של עבודה דרך טרמינל:

```
>> ./simCOVID config.dat first_generation.dat
```

1. מרכיבי עולם הנגיפים, אשר יוזנו לתכנית בתחילתה בקובץ הראשון (config.dat בדוגמא שלעיל), יתוארו בפורמט הבא: השורה הראשונה תכיל את מימד מחרוזות התווים  $\ell$ , השורה השנייה תכיל את מחרוזת המטרה  $\vec{r}$  (באמצעות  $\ell$  קואורדינטות מופרדות ברווחים), והשורה השלישית תכיל את מספר מעברי הזמן המקסימליים שעל התכנית לסמלץ. להלן דוגמא של קובץ קלט חוקי של תיאור עולם הנגיפים:

```
10
A A A T G C C G A A
10000
```

2. על קובץ הקלט השני (קובץ הדור ההתחלתי; first\_generation.dat בדוגמא שלעיל), להכיל בשורה הראשונה את מספר הנגיפים בעולם (גודל האוכלוסייה שיישאר קבוע במהלך הריצה), ובשורות לאחר מכן את סוגן ואת מחרוזות התווים ההתחלתיות של כל הנגיפים. כל שורה תתחיל בתו המגדיר את טיפוס הנגיף (מתוך שלשת התווים {'a', 'd', 'o'}), ולאחר מכן תכיל את המחרוזת ההתחלתית של הנגיף המתאים, באמצעות  $\ell$  קואורדינטות מופרדות ברווחים. להלן דוגמא של קובץ קלט חוקי של אתחול הנגיפים:

```
8
a A T T C C A G C T T
a T A A G G T C G A A
d A C T G G C T A A A
o T T T C G A A A G C
a A T A G G C T G C G
a A T T G G C T A A A
o T T T C T T A A G C
o A T A G G C T G C C
```

3. במקרה של קלט בלתי-חוקי (למשל, שגיאת קבצים, חוסר-עקביות במימדים וכיוצא באלה), התכנית תסיים את ריצתה בשלב זה ותדפיס הודעת שגיאה מתאימה לערוץ השגיאות הסטנדרטי.

4. בהנחה שהקלט של המשתמש חוקי, על התכנית להתחיל את מעברי הזמן בעולם, תוך עדכון מחרוזות הנגיפים מדור לדור. התכנית תרוץ עד שאחת המחרוזות תגיע לעוצמת התאמה 1.0, או עד שמספר המעברים המקסימליים יחלוף.

# נושאים מתקדמים בתכנות מונחה-עצמים, אביב 2022

## פורמט פלט

בסיום מעברי הזמן על התכנית להחזיר כפלט שרשור של שלושה תכנים דרך ערוץ הפלט הסטנדרטי, כאשר שורת רווח תפריד בין כל תוכן:

1. מחרוזות הנגיפים של הדור האחרון (בפורמט הקלט)
2. מניית המופעים של כל "מחרוזת אב קדמון" (בפורמט הקלט בצירוף מספר שלם למניית הרפרנסים, מופרד ברווח)
3. מחרוזת הנגיף החזק ביותר בהיסטוריית הריצה (בפורמט הקלט)

להלן פלט אפשרי עבור הדוגמא שלעיל –

```
a T T T C C A G C T T
a A A A G G T C G A A
d T G T G C G T A A A
o A A A C G A A A G C
a A T A G G C T G C G
a A T T G G C T A A A
o T A T C T T A A C C
o A A A G G C A G C C
```

```
a A T T C C A G C T T 2
a T A A G G T C G A A 0
d A C T G G C T A A A 1
o T T T C G A A A G C 1
a A T A G G C T G C G 0
a A T T G G C T A A A 2
o T T T C T T A A G C 1
o A T A G G C T G C C 1
```

```
d A A A T G C C G T T
```

# נושאים מתקדמים בתכנות מונחה-עצמים, אביב 2022

## דרישות מימוש

- אתם חופשיים לתכנן ולממש את התכנית כרצונכם, פרט לדרישות הבאות אותן עליכם לקיים במימוש שלכם:
- יש לממש מנגנון *reference counting* עבור זיכרון מחרוזת האב הקדמון. כלומר, כל מחרוזת תווים של הדור ההתחלתי תוחזק באמצעות מנגנון זה, אשר יעקוב אחר המספר העדכני של הנגיפים אשר מקורם במחרוזת זו. שימו לב שבעקבות הוצאת הנגיפים "החלשים" מהאוכלוסייה בכל דור, מחרוזות אב קדמון מסויימות עשויות להיעלם מן האוכלוסייה.
- מימוש *reference counting* יכול להסתמך על שקפי ההרצאה, או על כל רעיון אחר בו תבחרו, פשוט או מורכב כרצונכם.
- עליכם להשתמש בספריית STL עבור מבני הנתונים בפתרון שלכם.

## הנחות עבודה

- מספר מעברי הזמן (דורות סימולציה) הינו מספר טבעי: בהינתן ערך 0, התכנית תדפיס את מצב העולם ההתחלתי ותסיים; בהינתן ערך 1, התכנית תחשב מעבר דור בודד, תדפיס את מחרוזות הנגיפים ותסיים; וכך הלאה. על מספר הדורות לא לעלות על  $10^6$ .
- גודל האוכלוסייה הוא לפחות 2 נגיפים ולכל היותר 200 נגיפים.
- מימד מחרוזות תווים הוא לפחות 3 תווים; על המימד לא לעלות על 1000.

## דגשים

- אתם מצופים לתכנות בסטנדרטים הגבוהים של סיום הקורס "תכנות בשפת C++", ובפרט, לדיזיין מונחה-עצמים, שימוש מושכל בטכניקות העמסת אופרטורים, פולימורפיזם, ושימוש בתבניות במידת הצורך.
- במקרים של אי-תקינות קלטים, או במקרה של שגיאות זמן-אמת, יש להפעיל מנגנון חריגות בשכבות אובייקט הסימולציה ופניה, ולהציג למשתמש הודעות שגיאה מתאימות. בכל מקרה, מצב בו חריגה איננה נתפסת ומביאה לקריסת התכנית תיחשב לשגיאה בפתרון התרגיל.
- בתרגיל זה **אינכם** רשאים להשתמש במצביעים החכמים הסטנדרטיים של תקן השפה החדש C++0x: `unique_ptr`, `shared_ptr`, `weak_ptr`
- כאמור, אתם רשאים לעשות שימוש בחומרי ההרצאה והקוד שתואר בהם לצורך מימוש מנגנון *reference counting*.
- עליכם לוודא כי התכנית עוברת קומפילציית g++ התואמת את הקומפיילר שעל שרת המכללה ללא כל שגיאות או אזהרות כלשהן, ורצה בהצלחה.
- עליכם לתעד את הקוד באמצעות הערות המתארות בקצרה את הפונקציות השונות.
- מומלץ להשתמש בפונקציה `rand()` לצורך הגרלה אקראית המתפלגת באופן אחיד במימוש העדכונים ההסתברותיים.

## הגשה

- הכינו קובץ בשם `README.txt`. הכולל את שם ותעודת הזהות של הסטודנט/ית המגיש/ה; בקובץ זה אתם מוזמנים לכלול הערות ותיעוד כללי לגבי המימוש שלכם.
- עליכם להגיש במערכת Moodle קובץ ארכיב tar/zip/rar בשם `ex1.xxx`. המכיל את כל קבצי המקור (ממשק/מימוש) הנדרשים לקומפילציה, וכן את הקובץ `README.txt`.

אי-הקפדה על ההנחיות תגרור הורדה בציון התרגיל. לא תתקבלנה הגשות באיחור!